

Sensitivity of Numerical Weather Prediction Model Configuration on Aviation Weather Hazard Prediction

Annual Report of NCAR/RAL Model Development & Enhancement FY11 Efforts

compiled by Matthias Steiner

with sections contributed by

Joe Grim, Kyoko Ikeda, Frank McDonough, James Pinto,
Marcia Politovich, Bob Sharman, Jenny Sun, and Mei Xu

15 September 2011

“This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy or position of the FAA.”

Table of Content

1. Introduction	page 3
2. Case Selection and Numerical Experimentation	4
3. Analyses focused on Convection	8
3.1 Assessment of the 13 km and 3 km Baseline Simulations	8
3.2 Sensitivity of the Forecast to Model Physics Options	22
3.3 Assessment of real-time HRRR Forecasts of MCS during June – August 2010	28
3.4 Summary of convective Analyses and Suggestions for future Work	33
4. Analyses focused on Turbulence	36
4.1 Qualitative Assessment of Baseline and Sensitivity Runs	36
4.2 Quantitative Assessment of Baseline Runs	43
4.3 Summary of Turbulence Analyses and Suggestions for future Work	45
5. Analyses focused on In-flight Icing	47
5.1 Quantitative Analysis	47
5.2 Qualitative Analysis	52
5.3 Principal Findings	55
6. Analysis focused on cold-season Surface Precipitation Coverage and Phase	56
6.1 Introduction	56
6.2 Dataset	56
6.3 Analysis Results	58
6.4 Quantitative Analysis	65
6.5 Key Remarks and Recommendations	74
7. Summary of Findings and Outlook	76
7.1 Convection	76
7.2 Turbulence	77
7.3 In-flight Icing	78
7.4 Precipitation Type	78

1. Introduction

The FAA-supported model development and enhancement (MD&E) effort builds upon an ongoing, synergistic collaboration between the NOAA Earth System Research Laboratory (ESRL) Global Systems Division (GSD) and the NCAR Research Applications Laboratory (RAL). NCAR's expertise in numerical weather prediction (NWP) model development, data assimilation research, and weather analysis is utilized for carrying out a variety of sensitivity studies that are geared toward providing guidance for optimization of the NWP model configurations that are run operationally at NOAA/ESRL.

This year's focus on NCAR's side continued to take a broader look at the impact of particular model configurations on the detection and prediction of a variety of aviation hazards, including convective storms, turbulence, and in-flight icing. In addition, a careful evaluation of the model skill in the timing and location of predicting storms and the associated type of precipitation (i.e., rain, snow, or mixed phase) was conducted. Several major aviation impact days were selected so that the weather encountered on these days included convective storms, turbulence, and in-flight icing conditions. The purpose of that was to analyze identical days from different perspectives to document how changes in model configuration impact various aviation hazards. Most of the work in the past has been focused on convective storms, but optimization for convection may not necessarily yield a good configuration for turbulence and in-flight icing as well.

This report documents analyses carried out with a joint focus from convective and winter storms, turbulence, and in-flight icing on several major impact days. The results discussed in the report provide important feedback toward optimization of the NWP model configurations that are run operationally at NOAA/ESRL, especially the 13 km WRF Rapid Refresh (WRF-RR) and the 3 km High-Resolution Rapid Refresh (HRRR) models. In light of the ever evolving models, these jointly focused analyses need to be continued and expanded to achieve a solid basis from which to draw conclusive evidence on how to best configure the WRF-RR and HRRR models. Additional sensitivity studies are documented with a specific focus convective storm only.

Section 2 of the report documents the sensitivity analyses and the events selected. Sections 3, 4, 5, and 6 discuss the various results obtained from a convective storms, turbulence, in-flight icing, and precipitation type perspective, respectively. An overall summary and outlook is provided in section 7.

2. Case Selection and Numerical Experimentation

Seven cases were selected that all included episodes of convection, turbulence, and in-flight icing causing significant impact on aviation. Three of the cases occurred during the summer and early autumn (June – September 2010), one case was during late autumn (October 2010), and the other three were cold season cases (February 2010, November 2010 and February 2011). A summary of the phenomena of interests for those events is given in Table 2.1. Additional descriptions of these events are provided with the case analyses presented in subsequent chapters.

For each of the seven cases, 13 km and 3 km baseline simulations were conducted with the WRF-ARW model (Skamarock et al. 2008), emulating the NOAA/ESRL Rapid-Refresh (RR) model and High-Resolution Rapid-Refresh (HRRR) model. The model configurations for these forecasts are summarized in Table 2.2. The WRF model core used in the baseline simulations is based on GSD's modified version of WRFv3.2, which was the code used in the 2010 HRRR and RR operational runs. A few updates were applied to the GSD code, including the WRF v3.3.0 Thompson microphysics. The model parameters and options are similar to those of the operational RR and HRRR, with the exception of the 6th order numerical diffusion. A diffusion coefficient of 0.25 is used in the baseline simulations here, while the scheme is turned off in the current operational HRRR (since 2010).

The baseline simulations at two horizontal resolutions of 13 km (RR) and 3 km (HRRR) provide an opportunity to examine the influence of model horizontal resolution on convection, turbulence and icing forecasts. The 13 km model configuration uses a convective parameterization scheme whereas the 3 km model configuration resolves convection explicitly. As summarized in Table 2.1, additional sensitivity experiments were also conducted with different initial conditions (initialization time), physical parameterization schemes, numerical diffusion, and vertical grid resolution. The physical parameterization schemes include PBL schemes (YSU, MYJ, and QNSE), microphysical schemes (Thompson v3.3.0 and v3.3.1, Milbrandt-Yau, and WSM6), and LSM schemes (RUC and Noah). Due to computational limitations, only a subset of cases was selected for each type of sensitivity at 3 km model resolution.

Table 2.1: Summary of weather events and numerical experiments for the joint cases of convection, turbulence, in-flight icing and winter storm precipitation type studies.

<i>Date</i>	<i>Model Initialization Time (UTC)</i>	<i>Phenomena of Interest</i>	<i>Forecast-Sensitivity Experiments</i>
4 Feb 2010	12	Convection: Strong storms across the western Gulf states associated with warm moist southerly surface flow and upper level jet streak. Icing: Upper level trough over the Rockies forming clouds; overrunning clouds over the Central Plains; storms moving into Pacific NW and over Mississippi Valley.	<ul style="list-style-type: none"> • Horizontal resolution
17 Jun 2010	12	Convection: Scattered storms in SE. Icing: Low moving across northern Plains; strong cold front into northern and central Rockies with associated clouds and precipitation; storms across northern Plains.	<ul style="list-style-type: none"> • Horizontal resolution • PBL
19 Jul 2010	00, 03, 06	Convection: Multiple MCSs in Central Plains. Turbulence: Turbulence associated with storms over Central Plains.	<ul style="list-style-type: none"> • Horizontal resolution • PBL • MP
16 Sep 2010	12	Convection: Thunderstorms along cold front (see icing below) over the central Appalachians near synoptic warm front; some storms over western TX. Icing: Surface low moving across western Great Lakes forming clouds; cold front across Mississippi Valley and warm front over the southern Great Lakes to the Atlantic Coast.	<ul style="list-style-type: none"> • Horizontal resolution • MP
25 Oct 2010	12	Convection: Severe thunderstorms moving from the Mississippi Valley to the Southeast coast. Icing: Clouds along cold front moving into Plains; clouds associated with a stationary front over the Great Lakes; convective icing over Southeast. Turbulence: Very strong jet across NW.	<ul style="list-style-type: none"> • Horizontal resolution • PBL • MP • LSM

16 Nov 2010	12	<p>Convection: Severe thunderstorms along stationary/cold front over mid-Atlantic, Southeast and lower Mississippi Valley.</p> <p>Icing: Clouds associated with surface low over Ohio Valley and associated stationary/cold front; mountain clouds over the northern and central Rockies.</p> <p>Turbulence: Deep trough over Great Plains; several short-waves present; northwest flow aloft over western half of US, southwest flow over eastern half.</p>	<ul style="list-style-type: none"> • Horizontal resolution • PBL • 6th order diffusion • vertical resolution
1 Feb 2011	12	<p>Convection: Severe thunderstorms over the lower Mississippi Valley and Southeast in warm sector of cyclone ahead of strong cold front.</p> <p>Icing: Surface cyclone over the Midwest with very cold air moving south from Canada; warm front over Great Lakes and Northeast, cold front across the central and southern Plains.</p> <p>Turbulence: Deep trough over the Rockies and Plains; jet streak rounding base of through over TX, and moving south over the intermountain West.</p>	<ul style="list-style-type: none"> • Horizontal resolution

Table 2.2: WRF-ARW model configuration for the 13 km and 3 km baseline simulations.

<i>Model Parameter</i>	<i>“HRRR”</i>	<i>“RR”</i>
WRF-ARW model version	3.2 with NOAA/GSD code changes* + v3.2.1 PBL updates + v3.2.1 RRTMG updates + Thompson v3.3.0 mp	3.2 with NOAA/GSD code changes* + v3.2.1 PBL updates + v3.2.1 RRTMG updates + Thompson v3.3.0 mp
Horizontal grid	CONUS; Dx = 3 km 1680 × 1100 gridpoints	CONUS; Dx = 13 km 400 × 250 gridpoints
Vertical levels	51	51
Initialization	RUC	RUC
PBL scheme	MYJ	MYJ
Land-surface model	RUC	RUC
Microphysics scheme	Thompson	Thompson
Convective scheme	none	Grell-Devenyi
Radiation schemes	Dudhia (shortwave) RRTM (longwave)	Dudhia (shortwave) RRTM (longwave)
6th order horizontal diffusion coefficient	0.25	0.25

(*) For real-time HRRR forecasts in summer 2010, NOAA/GSD implemented minor changes to the WRF3.2 code involving the RUC land-surface model, diagnostic output fields, and digital filter initialization.

References for Section 2

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: *A description of the Advanced Research WRF version 3*. NCAR Tech. Note NCAR/TN-475+STR, 113 pp.

3. Analyses focused on Convection

Performance and characteristics of the 13 km (RR) and 3 km (HRRR) baseline simulations, and associated sensitivity simulations, are discussed in this chapter in terms of their prediction of precipitation and other convection related features. Qualitative evaluations as well as quantitative assessments are presented. While all of the seven cases are included in the discussion, an emphasis is given to the four warm-season cases and the associated sensitivity analysis. Also included in this chapter is a performance assessment of the real-time HRRR (run at NOAA/ESRL) in predicting mesoscale convective system initiation during the summer of 2010.

3.1 Assessment of the 13 km and 3 km Baseline Simulations

3.1.1 Qualitative Assessment

The simulated storm type, location, intensity, and timing are examined and compared with radar observations, satellite images, objective analyses and Stage IV (raingauge adjusted radar) estimates. Given that most of the forecasts under evaluation are initialized at 12 UTC, the convective features of interests are usually most intense in the 6 – 12 hour forecast range.

The model forecast skills are clearly case-dependent and feature-dependent. In general, the 3 km simulation performs reasonably well on features of relatively large spatial and temporal extent, such as lines of convection, regional scattered convection, and winter storms. The model usually places these weather systems in the approximate region as observed, depicts storm structures with realism, and predicts the convective mode correctly. On the other hand, errors are evident in the predicted mesoscale details, such as in the exact location and orientation of line storms and storm clusters, initiation and maintenance of MCSs, and timing and strength of sea breeze convection. Such errors can be significant even in very short-term, say 1 – 3 hour forecasts.

The 13 km simulations are able to pick up most of the major precipitation systems as well. However, with nearly all of the precipitation produced through the convective parameterization scheme rather than explicitly through model dynamics and microphysics, the 13 km forecasts tend to only predict regions of precipitation, but lack detailed information on storm structures and organizations. Another obvious problem with the 13 km is over-prediction, especially in the warm-season cases (more on this will be discussed in Section 3.1.2). The model also seems to experience a “spin-down” period in the first hour. Large areas of false-alarm precipitation are predicted immediately following the initialization, especially over the ocean in the southeast part of the domain. Recall that both the 13 km and 3 km baseline simulations are initialized with a RUC analysis that has incorporated radar reflectivity assimilation. The problem seems to indicate an incompatibility between the initial conditions and the convective parameterization scheme used in the 13 km simulations. A possibility is that the RUC analysis was tuned for RUC model to produce sufficient convection and may be too hot for WRF (Ming Hu, personal communications).

On 17 June 2010, scattered convection started in the southeast around noontime and lasted through late afternoon. The 3 km baseline simulation captured the initiation of storms and convective mode in the region (Fig. 3.1). The 13 km forecast predicts swaths of precipitation and suffers from severe over-prediction. Both 3 km and 13 km forecasts have false alarms in Iowa and southeastern Nebraska.

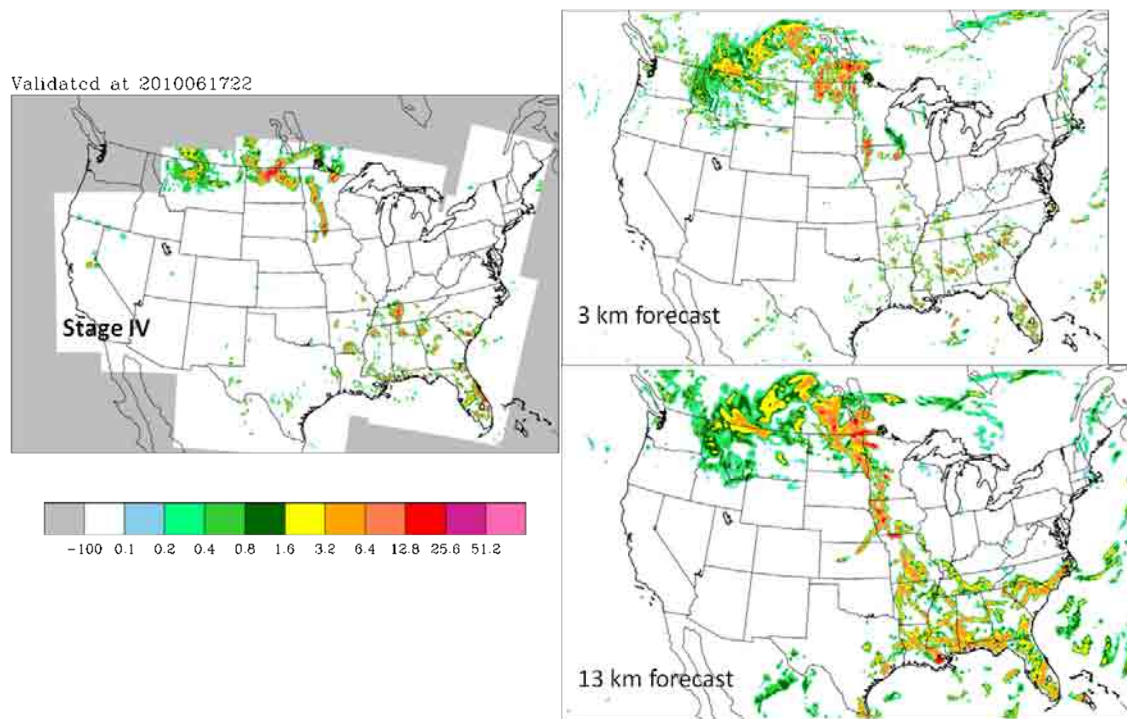


Figure 3.1: Observed (left) and forecast 1-hour rainfall from 3 km (upper right) and 13 km (lower right) model runs. The valid time is 22 UTC of 17 June 2010, 10 hours after the model initialization.

The weather of interest on 19 July 2010 is the nocturnal MCSs development in the Central Plain north of a stationary front (Fig. 3.2). The system in the southwestern part of South Dakota was a re-development around 06 UTC on the southern part of a dissipating system advected from the west. A second system, to the southeast at the boundaries of South Dakota-Nebraska-Iowa, initiated at 07 UTC and quickly grew into an MCS. Most of the real-time HRRR forecasts valid at 07 UTC of the day missed the MCS initiation (ref. Section 3.3). The 13 km and 3 km baseline simulations were initialized at 00 UTC. The simulations do not predict the timing and locations of the MCSs well (Fig. 3.2 upper-left).

Two more 3 km simulations using the same model parameters as in the baseline, yet with a smaller domain centered on the Central Plains, and initialized at 03 UTC and 06 UTC respectively, have also been conducted. With shorter lead time, these two simulations (Fig. 3.2 lower panels) show improved forecasts of the location of the system in SD, but they do not predict the up-scale growth and new initiations well. Even with radar data enhancement in the model initial conditions, predicting MCS initiation and maintenance continues to be a challenge to high-resolution models like the HRRR. On the other hand, the vast differences among the forecasts with different initialization time also reflect that the forecast is very sensitive to model initial conditions.

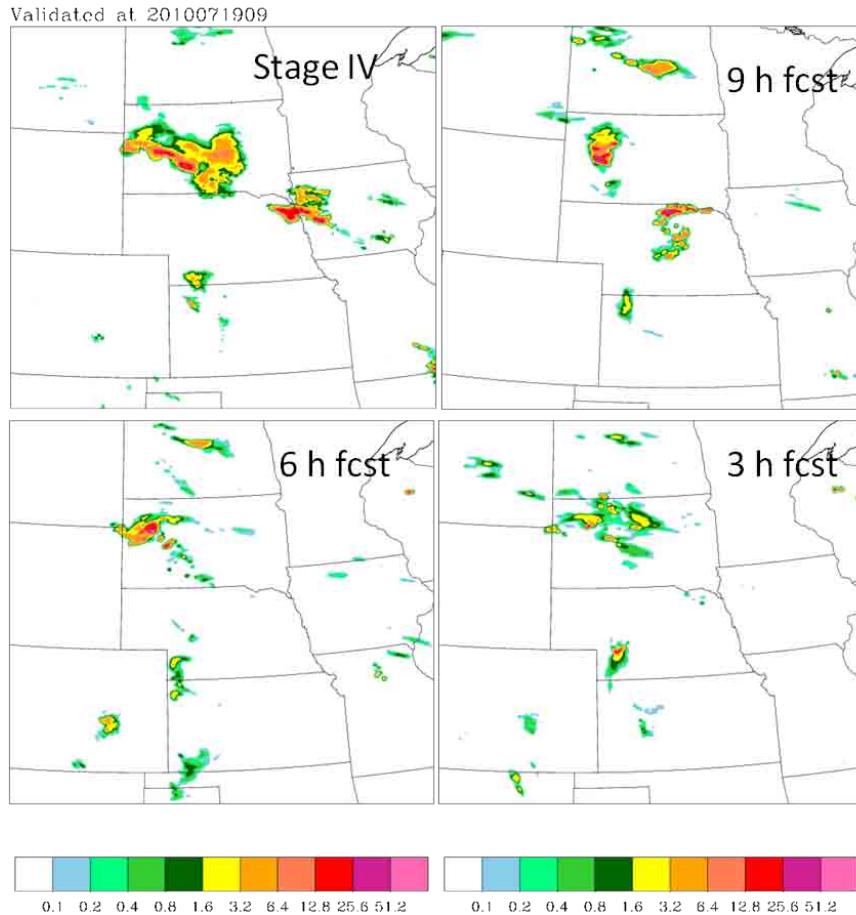


Figure 3.2: Observed (upper left) and forecast 1-hour rainfall from 3 km model runs initialized at 00, 03, 06 UTC on 19 July 2010. The valid time is 09 UTC of 19 July 2010.

The 16 September 2010 was a CoSPA blitz day. The convective features of interest were the long-lived, intensive MCS moving through the Midwest and the late afternoon severe storm outburst passing over New York City. The latter system lasted only for a few hours, but it highly impacted the NYC airports. It was embedded in a large area of stratiform precipitation and then intensified quickly. Both the 3 km and 13 km baseline simulations successfully predicted this system with small errors (Fig. 3.3). For the long-lived MCS in Midwest, the forecast storms are too weak and wide spread.

In the afternoon of 25 October 2010, two areas of convective activities are of interests, the weakly forced convection in NE and convection in the Carolinas and off the coast (Fig. 3.4). Although the 3 km and 13 km forecasts predict convections in both regions, the errors are relatively large. In the forecasts, the SE convection moves too much north and dissipates too early, and in the NE, the predicted convection initiates, intensifies and dissipates too late. In the 6-hour forecasts shown in Fig. 3.4, the predicted NE convection is too strong, while there is no convection predicted in the Carolinas.

Validated at 2010091700

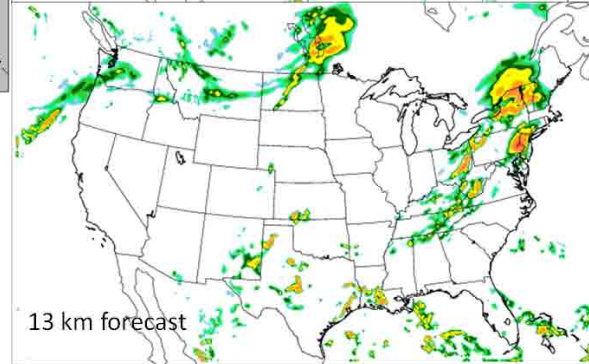
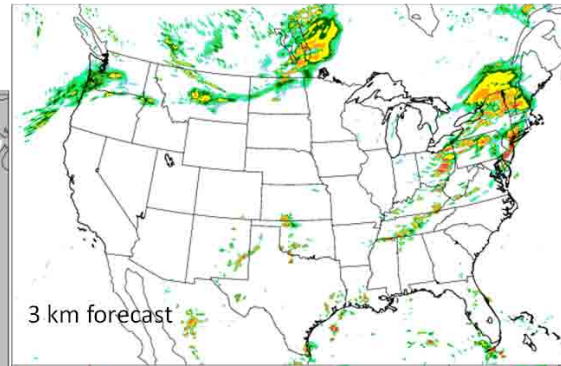
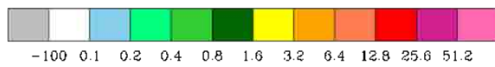
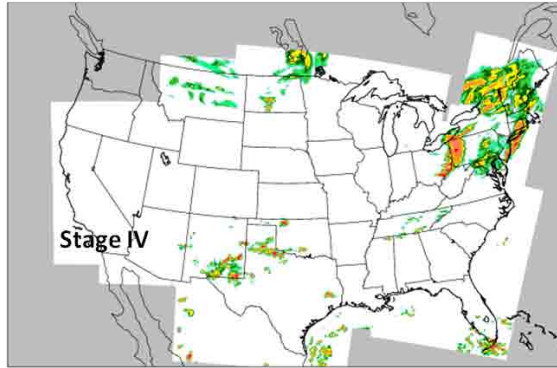


Figure 3.3: Same as Fig. 3.1 but for 00 UTC of 17 September 2010, 12 h after model initialization.

Validated at 2010102518

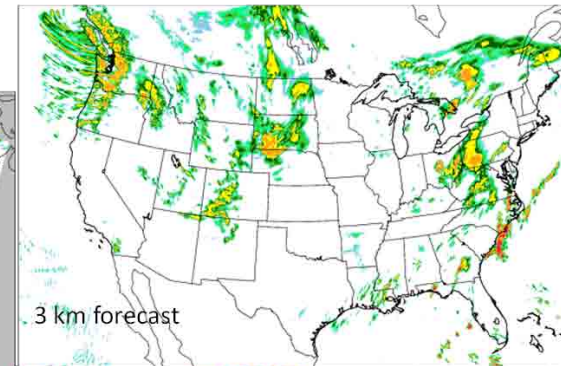
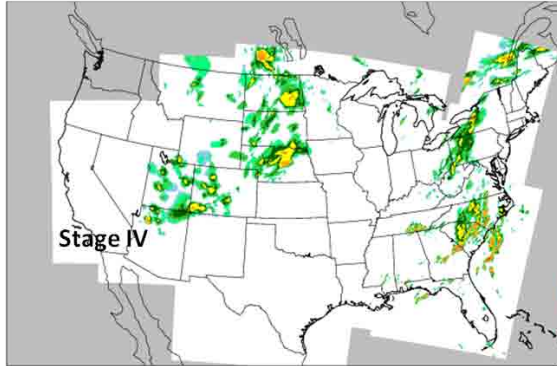


Figure 3.4: Same as Fig. 3.1 but for 18 UTC of 25 October 2010, 6 hours after model initialization.

3.1.2 Quantitative Rainfall Forecast Skills

An objective verification of the predicted precipitation is performed for both baseline simulations using the Stage IV estimates (Lin and Mitchell, 2005) as truth. To verify the precipitation forecasts, Stage IV hourly rainfall rates were first interpolated to the 3 km and 13 km model grids respectively. Skill scores were then calculated by comparing forecasts against observations at model grid points. Fractional Skill Scores (FSS) (Roberts, 2005) for 1-hour precipitation accumulations were calculated for a few selected precipitation intensity thresholds (0.1, 1.0 and 2.5 mm/h) using a radius of 50 km (Fig. 3.5) and 100 km (not shown) respectively. Also discussed here are the domain average rainfall rates from the forecasts and observations (Fig 3.6). The skill scores and average rainfall rates are calculated only for the areas where Stage IV estimates are available.

It is immediately noticeable that the precipitation forecasting skills for cold season cases are significantly better than those for warm season cases, for both the 3 km and 13 km runs. Of the four warm season cases, the 16 September 2010 case, which is associated with stronger synoptic forcing, has higher skill scores than the other three cases. For lighter precipitation threshold (0.1 mm/h), the skill scores are rather flat as a function of forecast lead time, while for heavier precipitation (2.5 mm/h), the skill scores drop after the first 2 hours. Another notable observation is that the 3 km simulations have higher scores than the 13 km simulation for almost all forecast hours and all situations. The advantages of the 3 km simulation are more significant for the warm-season events.

A comparison of the forecast and observed domain average rainfall shows significant over-estimation of the hourly rainfall rate for all cases but one (i.e., 19 July 2010) in both the 3 km and 13 km simulations. The over-estimation is more pronounced in the 13 km simulations than in the 3 km simulations for the warm-season cases (except for 19 July 2010). In the 3 km simulations for most cases (except for 4 February 2010), the model rainfall experiences a rapid growth in the first 1 – 2 hours after model initialization. On the other hand, in the 13 km simulations of warm-season cases, there is a large amount of excessive rainfall in the first 2 hours after model initialization, indicating a problem of overly hot start-up.

The 19 July 2010 simulations did not predict the timing and locations of the MCSs over the Central Plains well. The skill scores for the CONUS domain simulations, as well as those for the two 3 km simulations initialized at 03 and 06 UTC over the Central Plains domain, are relatively low. The domain average rainfall for the two simulations over the Central Plains shows under-estimations and delayed rainfall maxima (Fig. 3.7). The forecast timing of the onset and mature stage of the major MCSs was delayed by about 2 hours compared to the observation.

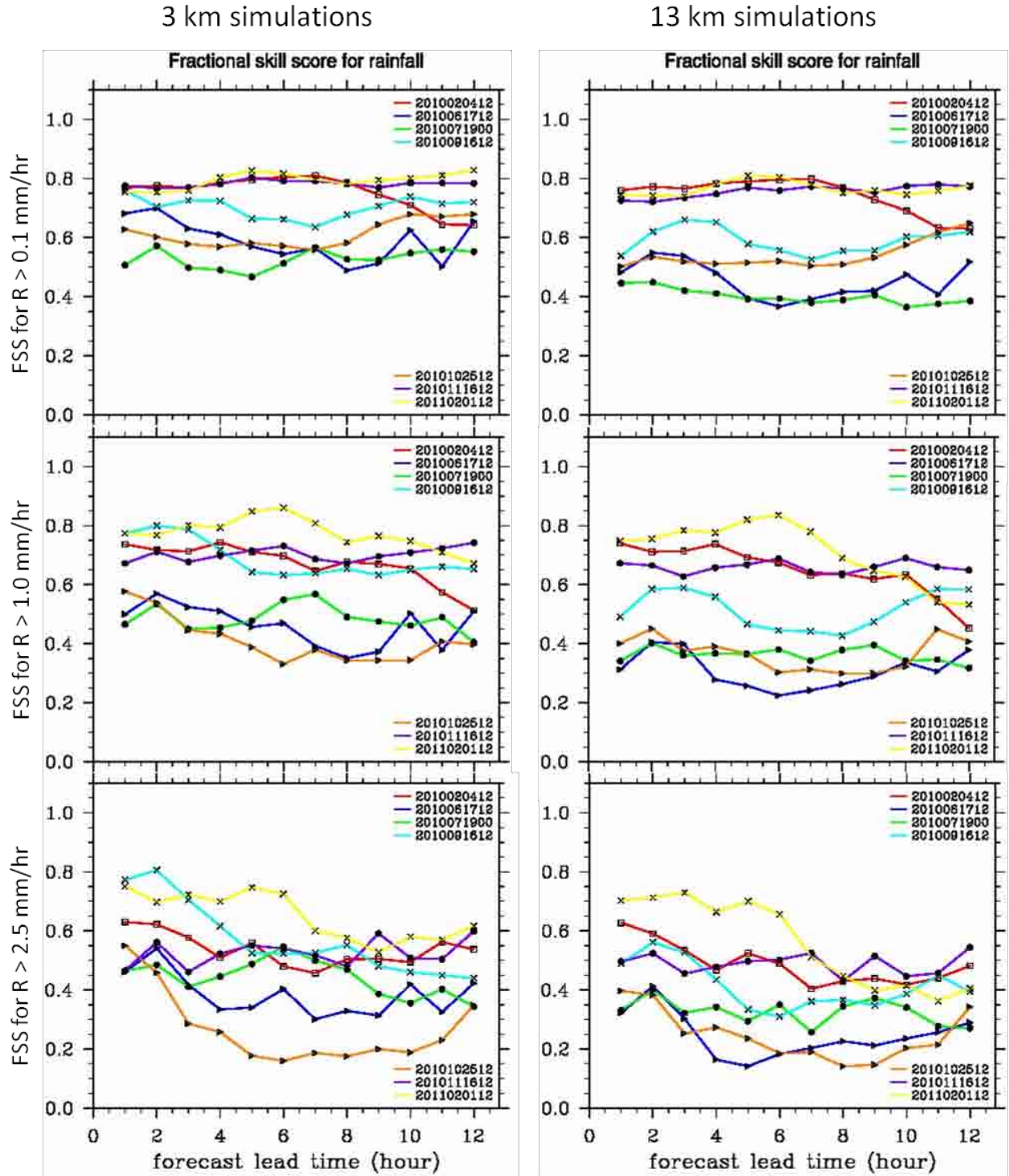


Figure 3.5: Fractional Skill Scores of the 3 km (left) and 13 km (right) baseline simulations for 0.1, 1.0 and 2.5 mm/h precipitation thresholds. Shown are skills for all seven cases.

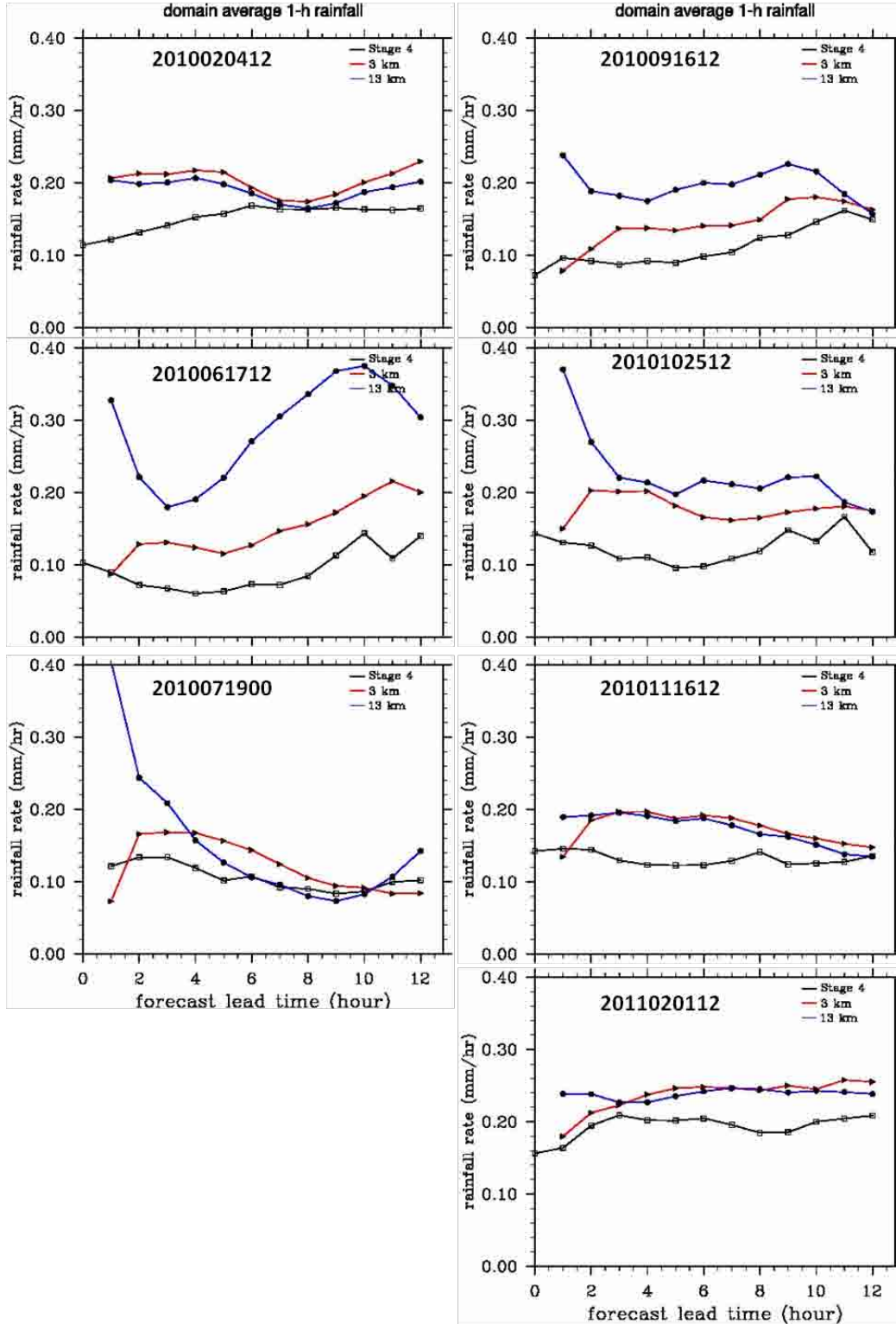


Figure 3.6: Domain average one-hour rainfall rate from Stage IV (black), 3 km (red) and 13 km (blue) baseline simulations over the CONUS domain for all seven cases.

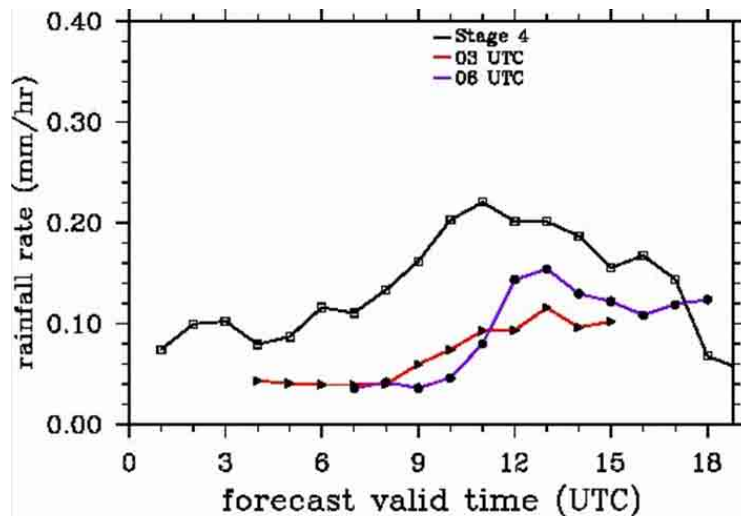


Figure 3.7: Domain-average hourly rainfall from Stage IV (black), and 3 km simulations over Central Plains initialized at 03 UTC (red) and 06 UTC (purple) of 19 July 2010.

3.1.3 Verification against Surface Station Observations

Surface station observations of 2 m temperature, humidity and 10 m wind were compiled from MADIS datasets for the selected events. Preliminary data quality control was performed to eliminate outliers and redundant data. Verification statistics were obtained by bi-linearly interpolating the modeled surface fields to the observation station locations, and comparing the modeled and observed variables using observations within ± 10 min from the model output time (i.e., hour). Due to a MADIS data gap in the local archive, surface verification statistics were not obtained for the case of 16 November 2010. The bias and root-mean-square errors (RMSE) in the surface fields of the 3 km baseline simulations are shown in Figs. 3.8a – c.

The 2 m temperature and moisture verification error statistics (Fig. 3.8a) show a cold and wet bias for all five CONUS forecasts initialized at 12 UTC. For the warm season cases, the temperature bias reaches (negative) maxima in the 2 – 5 hour forecasts, and then decreases with forecast time. For the cold season cases, the temperature bias becomes increasingly larger with forecast time. The bias and RMSE in the water vapor mixing ratio is clearly correlated with the total moisture amount, with larger errors in summer and smaller errors in winter. While the rainfall forecasts of the cold season simulations verify better than the warm season cases, this doesn't seem to be true for the surface fields.

All simulations over-predict the 10 m wind speed by a magnitude of about 1.6 – 3.6 m/s (Fig. 3.8b). The wind speed bias shows a rapid error increase in the first hour after initialization, and a relative flat error curve between 1 – 12 hours. This is a reflection of the start-up adjustment of the model wind, and it may be a result of a quick loss of the wind information incorporated by the data assimilation. The bias in wind direction is generally small, while the RMSE shows a 50 – 60° directional error. The two summer cases (17 June and 16 September 2010) have higher directional errors (RMSE). Interestingly, the wind direction RMSE is large already in the model initial conditions, but does not increase with forecast lead time. In two of the cases, there is actually some decrease with forecast length.

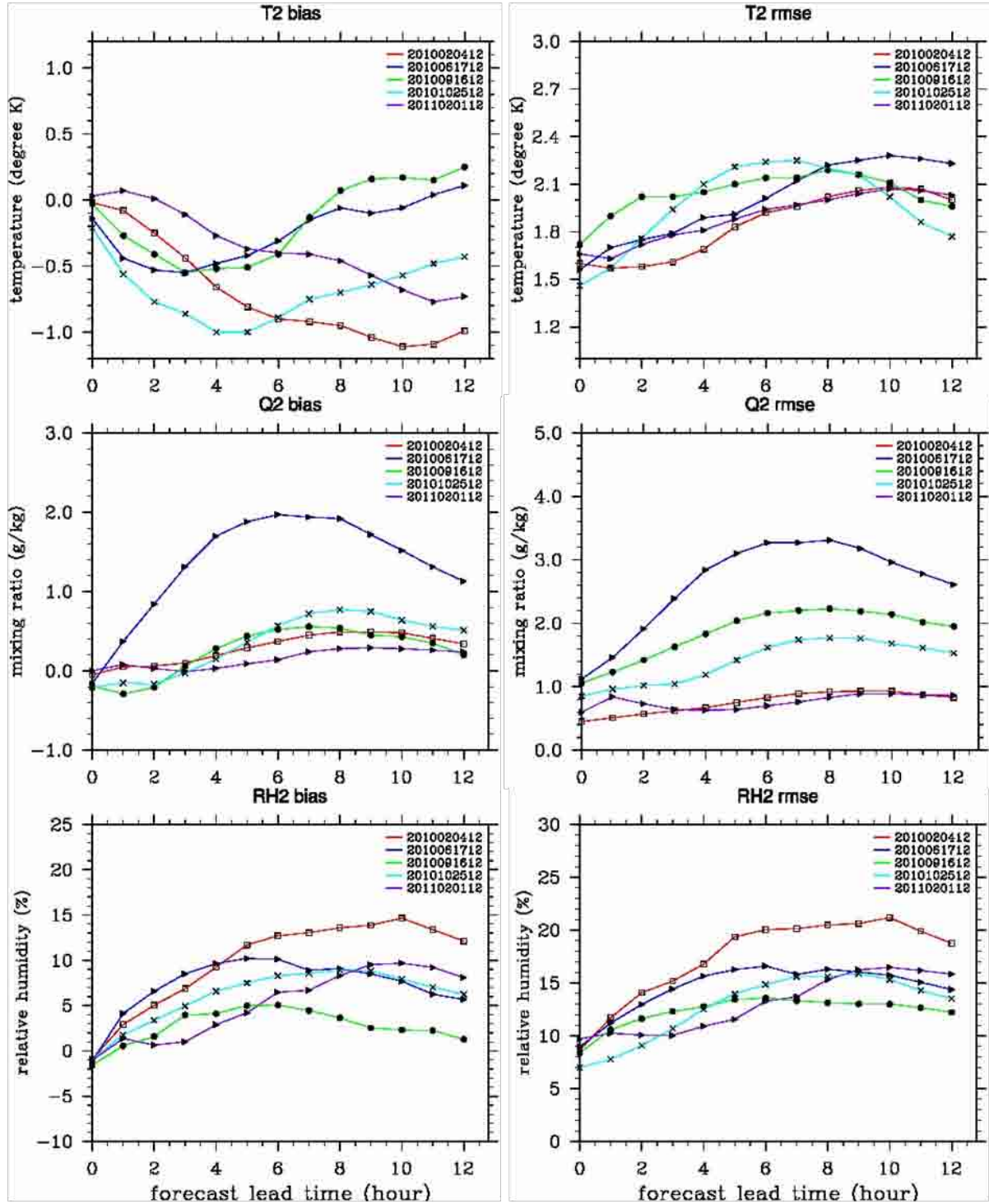


Figure 3.8a: Surface station verification statistics (bias and RMSE) of 2 m temperature and moisture for the 3 km baseline simulations for 5 cases.

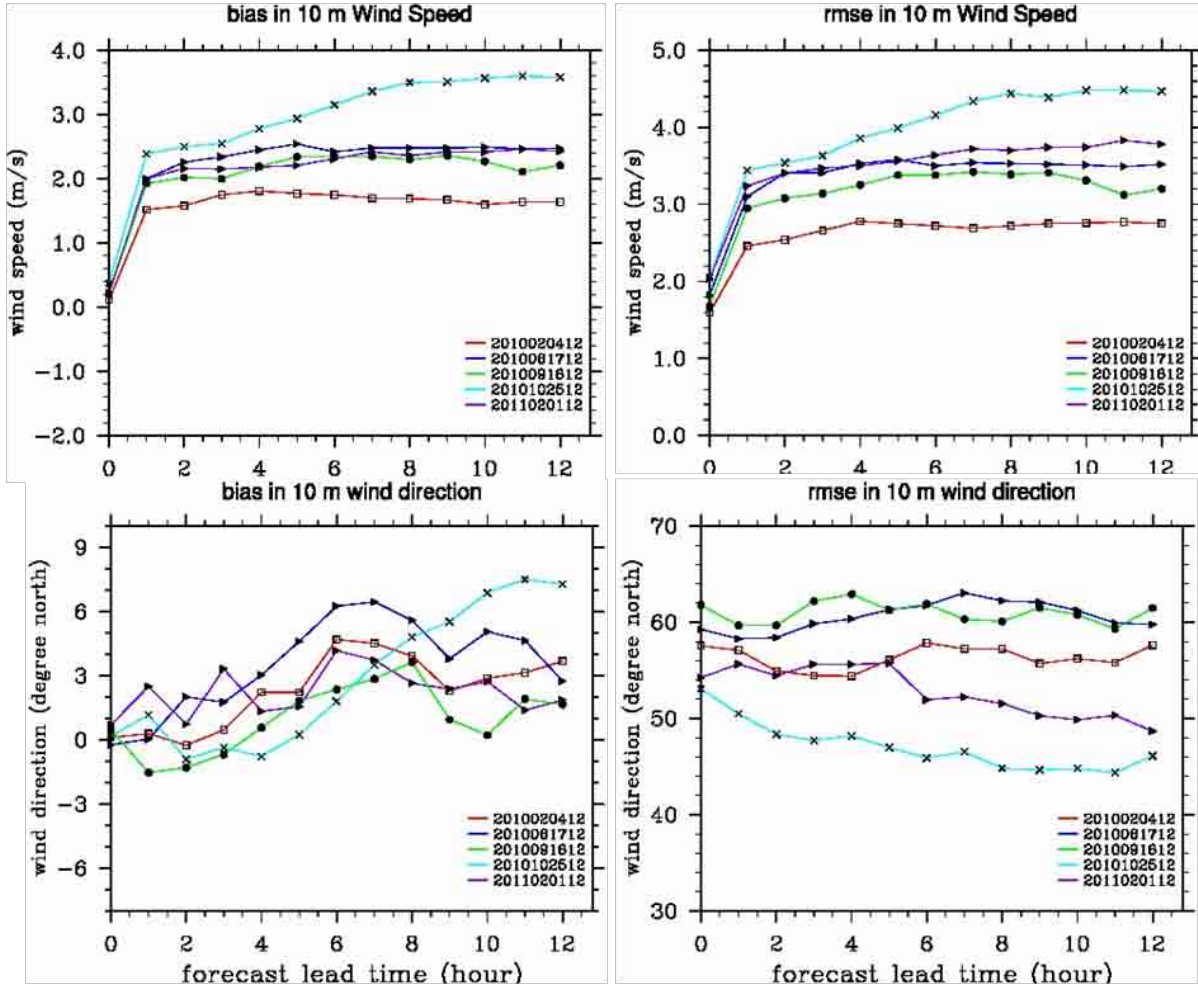


Figure 3.8b: Surface station verification statistics (bias and RMSE) of 10 m wind speed and direction for the 3 km baseline simulations for 5 cases.

Surface verification statistics for the case of 19 July 2010 is shown separately in Fig. 3.8c. Very different error properties are seen in this night to morning case. In the CONUS run initialized at 00 UTC, the simulated temperature has a warm bias during most of the forecast hours. Both the moisture and wind speed errors grow rapidly in the first hour following model startup. In the two simulations focused on the Central Plains, errors of all variables are growing rapidly during 13 – 18 UTC. The model seems to have difficulties in catching up with the night-to-morning transition and upscale growth of the MCS. Comparing the two forecasts starting at 03 and 06 UTC, it is found that while the shorter forecast (i.e., 06 UTC, green curves) has smaller errors for most of the verification hours, the differences are usually small, except at model initial time (06 UTC). Even though the new initial condition at 06 UTC verifies significantly better than the 3 hour forecast from 03 UTC, errors in the new forecast quickly grow to similar levels as in the older forecast in a couple of hours. This is especially true for the 10 m wind forecast.

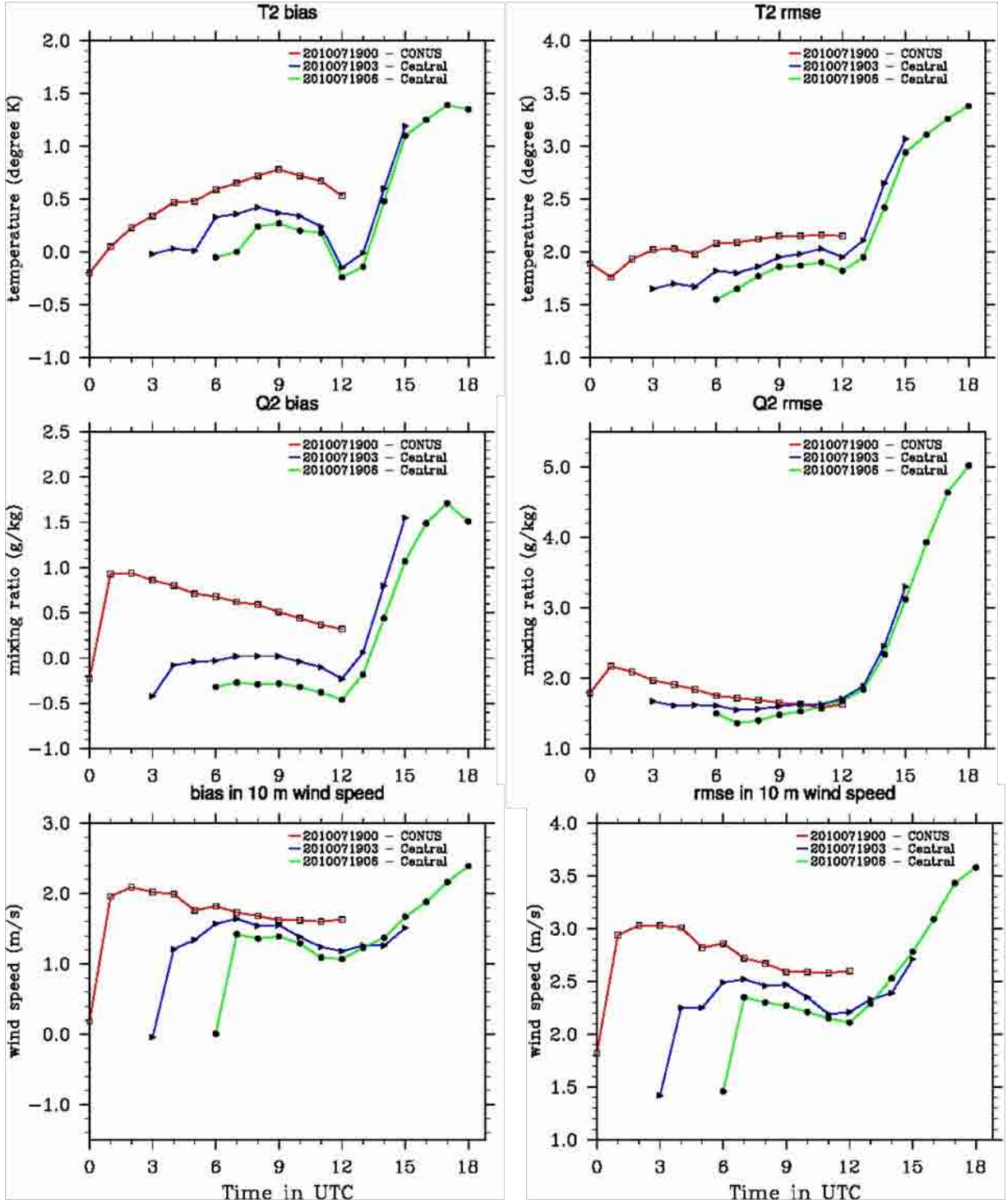


Figure 3.8c: Surface station verification statistics of 2 m temperature and moisture and 10 m wind speed for the 3 km simulations of the 19 July 2010 case.

The surface observations were also used to verify the location and depth of cold pools in the forecasts. An example is shown in Fig. 3.9 for the case of 16 – 17 September 2010. The general location of the cold pool behind the New York City outburst at 00 UTC on 17 September 2010 is

well predicted by the model, as indicated by the 2 m temperature and the temperature at the lowest model level (approximately at 8 m). The outflow boundary location and orientation compares well with the observations. However, the predicted cold pool is weaker and smaller than the observed. In addition, the temperature gradient across the outflow boundary is not as sharp in the simulation as in the observations. The observed zone of temperature gradient at the outflow boundary is narrow and compact, while the simulated gradient zone is wider. A similar situation is also seen in our analysis of the cross-boundary temperature gradient zone in the case of 19 July 2010 (not shown).

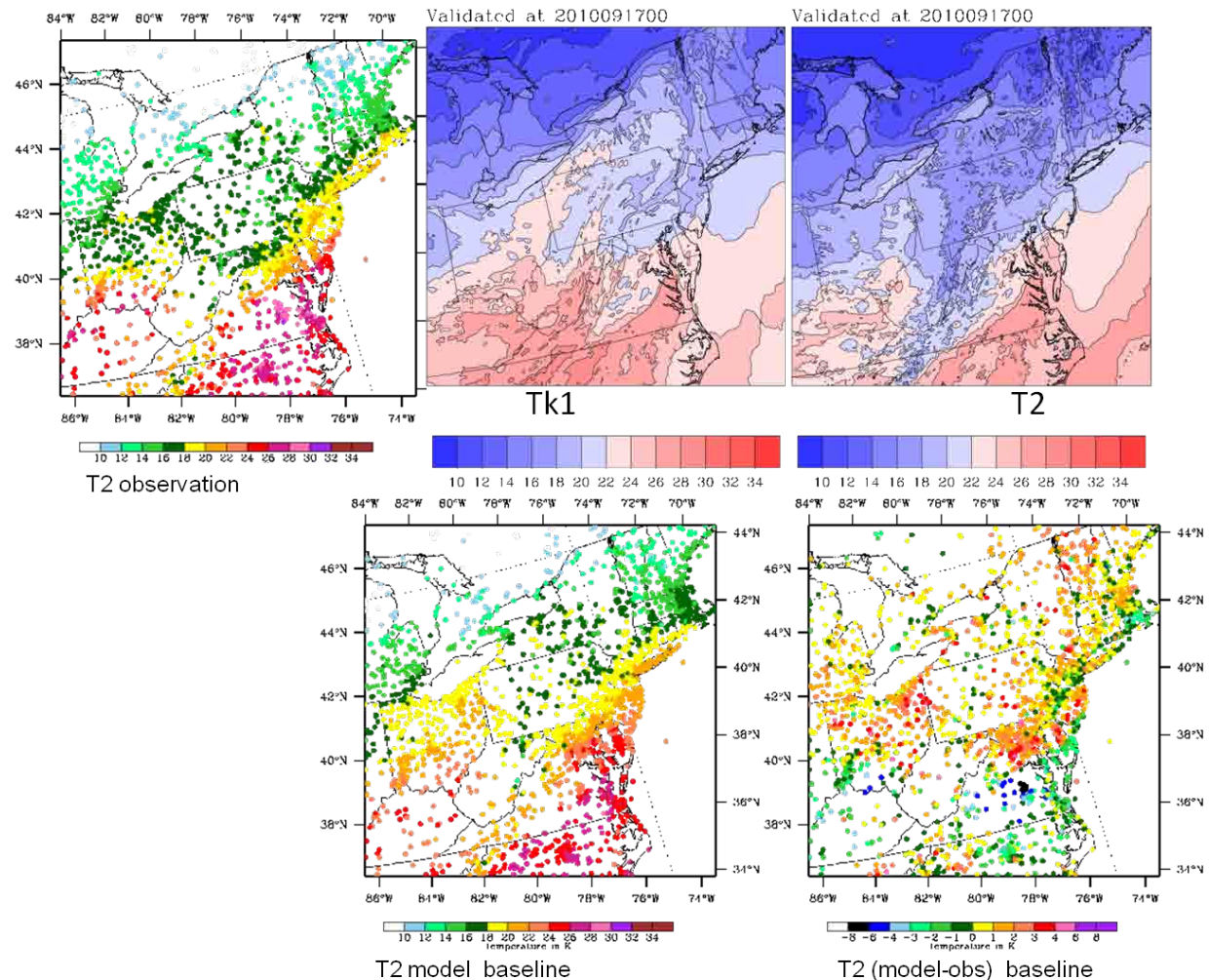


Figure 3.9: Observed 2 m temperature (T2 observation), modeled lowest-level (~ 8 m) temperature (Tk1), model derived 2 m temperature field (T2), model 2 m temperature at observation sites (T2 model baseline), and discrepancies between the model and observed 2 m temperature (T2 (model-obs) baseline), valid at 00 UTC of 17 September 2010 (12 hour forecast time).

3.1.4 Verification against upper-air Soundings

Upper-air fields and vertical wind shear from the model simulations were verified against profiler and rawinsondes sounding observations. The model variables were interpolated horizontally to the sounding locations, then both model and observations were interpolated vertically to a set of 40 pressure levels spaced 25 mb apart. There are approximately 130 – 150 distinct soundings at 00 and 12 UTC over the CONUS model domain. At hours other than 00 and 12 UTC, there are about 60, mostly wind profiles only. The verification statistics shown in Figs. 3.10a and b are the averages of the entire 0 – 12 hour forecast period.

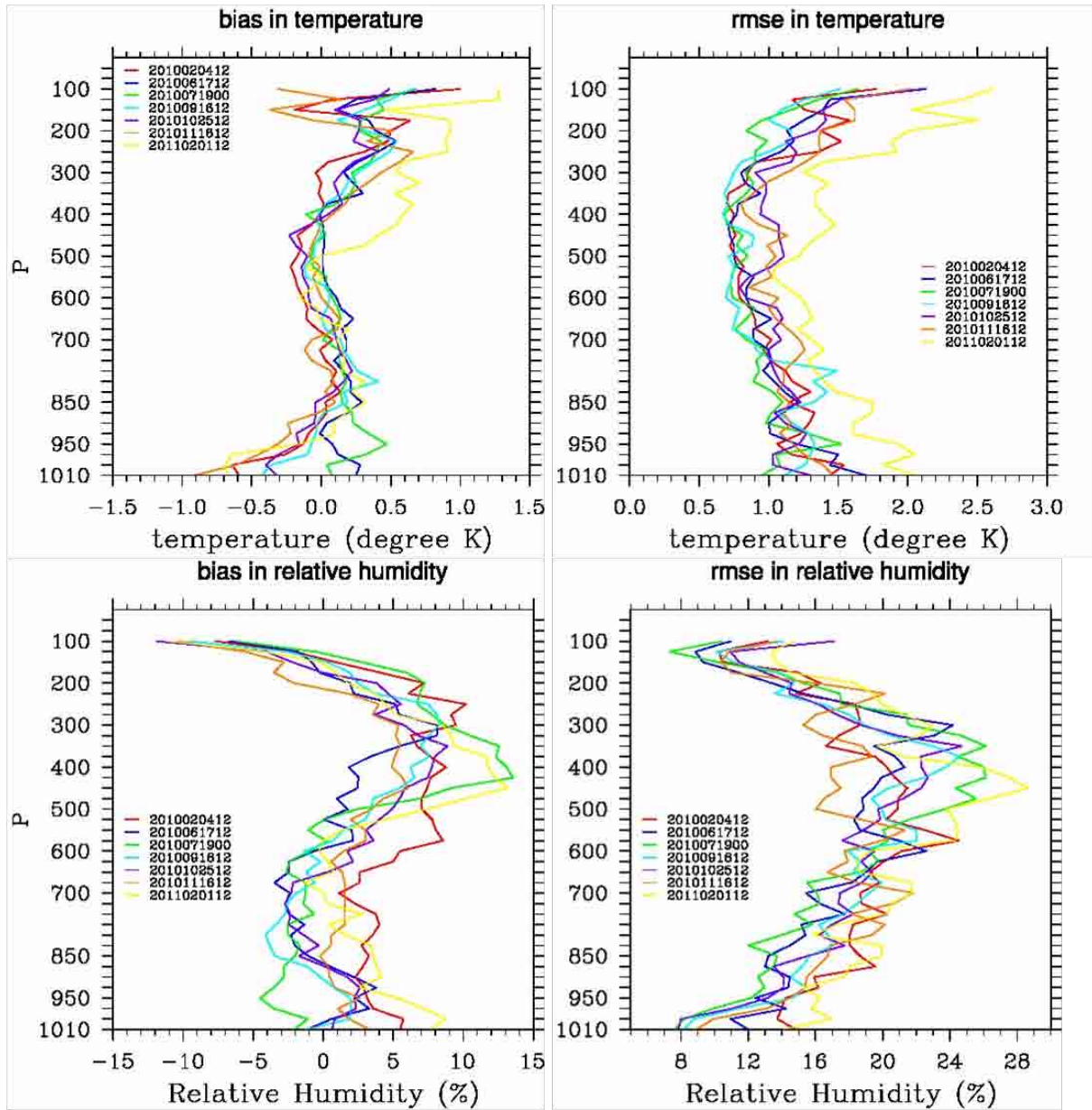


Figure 3.10a: Upper-air verification statistics of temperature and relative humidity for the 3 km baseline simulations.

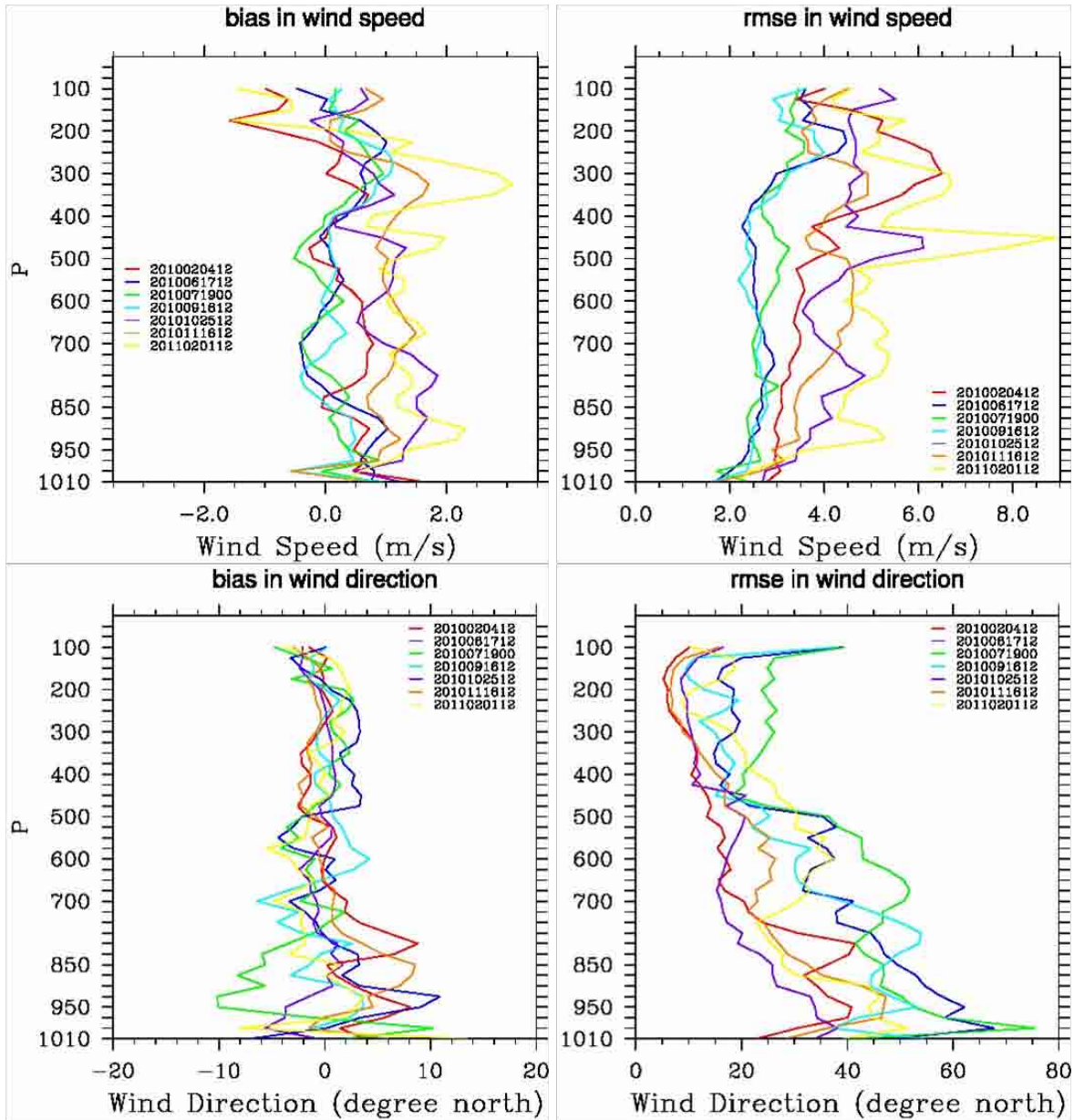


Figure 3.10b: Upper-air verification statistics of wind speed and direction for the 3 km baseline simulations.

The temperature errors (bias and RMSE) are relatively small at middle to upper levels (850 – 300 mb), with RMSE of about 1 m/s. An exception may be the 1 February 2011 case for which the RMSE in temperature is about 1.2 – 1.7 m/s. In general, the temperature errors in the colder, winter storm cases are somewhat larger. Larger cold biases in the boundary layer are shown in most of the seven cases, which is consistent with the above verification against the surface observations (section 3.1.3). The error in relative humidity increases with height in the middle atmosphere and has a maximum around 500 – 400 mb. Forecasts for the cold season cases tend

to have a wet bias from the surface throughout 200 mb, while for the warm season, the humidity bias is small or on the dry side at 850 – 500 mb. The large temperature bias in the low levels and large humidity RMSE throughout the troposphere might be two important factors that impact the quality of convective forecasting and requires further study.

There is a clear disparity between the 3 warm season cases (June – September) and the rest of the forecasts in terms of errors in wind speed and direction. The warm season cases have smaller errors in wind speed, but larger errors in wind direction. Apparently, with stronger wind aloft and more uniform wind directions in the winter, the model is able to predict the directions better. Both the wind direction bias and RMSE increase as the height decreases, which points to a deficiency of the current model in the low-level wind forecast. An accurate and dynamically balanced low-level wind analysis may improve the low-level wind forecast and hence the precipitation forecast.

3.2 Sensitivity of the Forecast to Model Physics Options

3.2.1 Forecast Sensitivity to the PBL Scheme

In the MD&E report for FY2010, we discussed forecast sensitivities to WRF planetary boundary layer (PBL) parameterizations based on 13 km forecast experiments. As an extension, in FY 2011, additional numerical experiments were conducted to study sensitivity of the 3 km forecasts to PBL schemes. In addition to Mellow-Yamada-Janjic (MYJ) that is used in the real-time HRRR and RR, and our baseline simulations, two additional PBL parameterizations, namely Yonsei University (YSU) and Quasi-Normal Scale Elimination (QNSE), were tested in two case studies (i.e., 17 June 2010 and 19 July 2010).

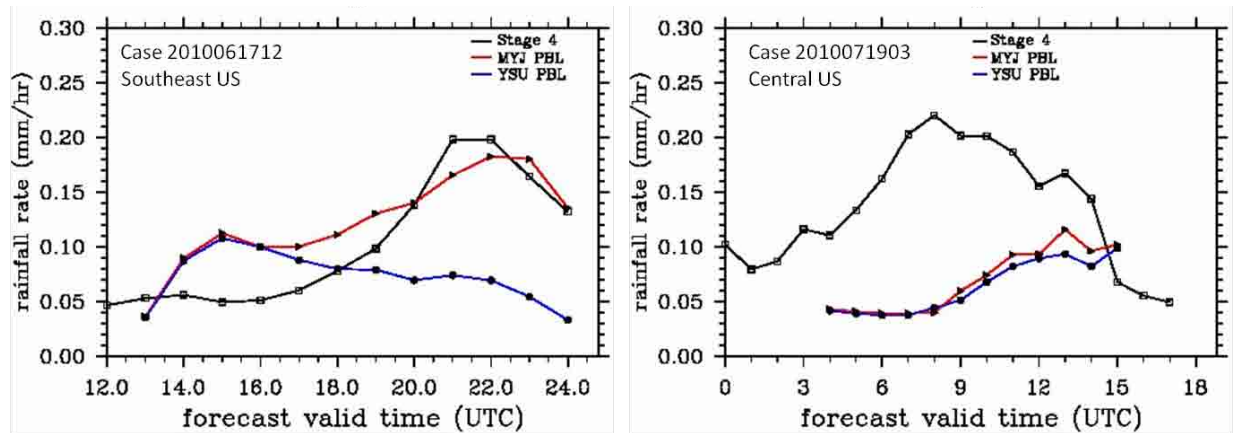


Figure 3.11: Area-average hourly rainfall from Stage IV (black), and 3 km simulations using MYJ (red) and YSU PBL (blue) schemes. Left: For forecasts initialized at 12 UTC on 17 June 2010 averaged over southeastern U.S. region. Right: For forecasts initialized at 03 UTC of 19 July 2010 averaged over the Central Plains domain.

For the phenomena of interests in the two cases, the differences caused by using QNSE (Sukoriansky et al. 2005) instead of MYJ PBL (Janjic 1994, 2001) appear to be small. In the

following discussions, our focus is on the YSU (Hong et al. 2006) scheme that is, apart from MYJ, another commonly used PBL scheme for WRF mesoscale and convective-scale forecasting. For both cases, the forecasts using the YSU PBL produced significantly less precipitation amount than the baseline runs (Figs. 3.11 & 3.12). The objective skills for rainfall show much smaller differences, however. Also, visual inspection suggests that the somewhat higher scores of MYJ may be due to over prediction and thus higher chance of hitting an observation. The FSS for the 17 June 2010 case is virtually unchanged (over CONUS) and the FSS for the 19 July 2010 is slightly degraded (not shown). Recall that the 19 July 2010 case focuses on MCSs that are already under-forecast in the baseline run.

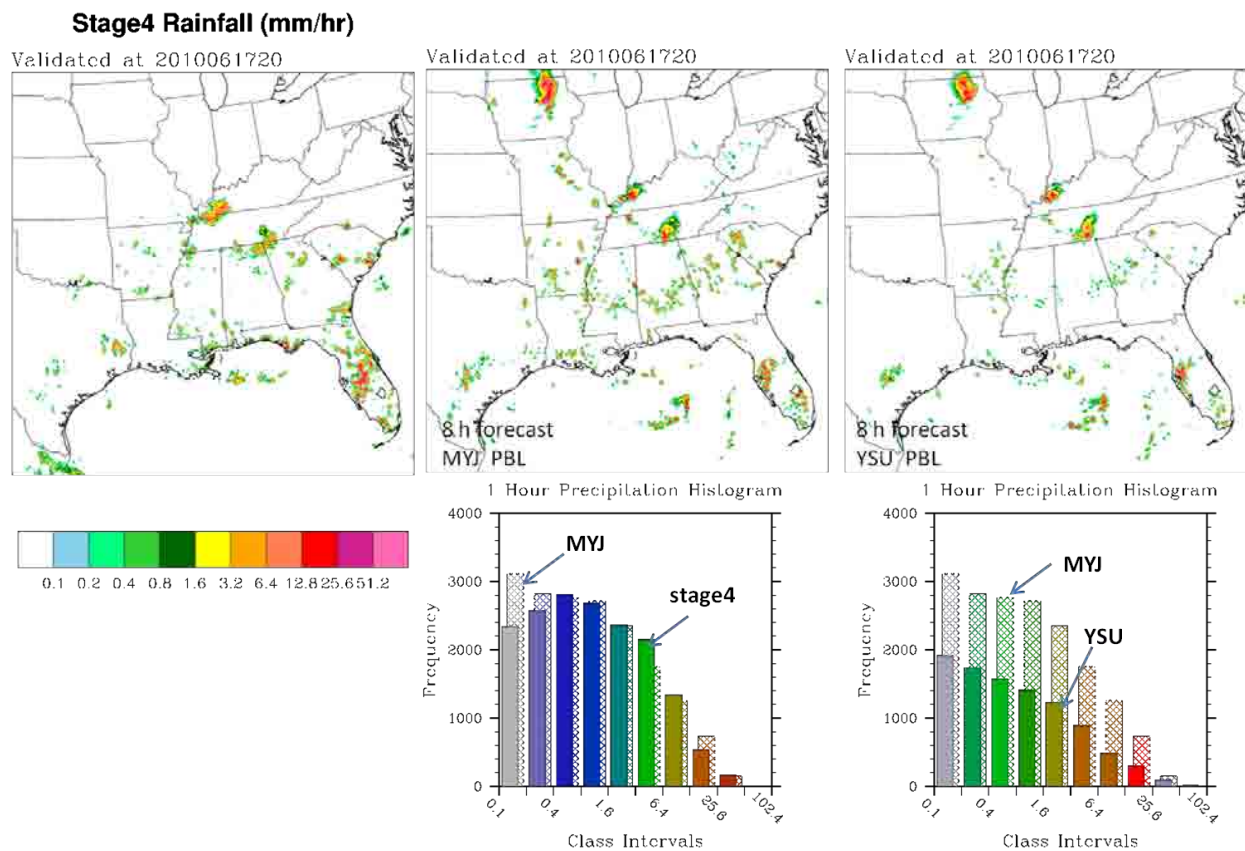


Figure 3.12: One- hour rainfall from Stage IV (left), and 3 km simulations using MYJ (center) and YSU PBL (right) schemes. The forecasts are initialized at 12 UTC and valid at 20 UTC on 17 June 2010. Also plotted are the histograms showing the number of grid points having precipitation rates in the specified intervals.

Surface station verification reveals that the surface fields in the MYJ forecast tend to be colder and wetter, while the YSU forecast is warmer and drier, especially during daytime. For the 17 June 2010 case, during late afternoon in some regions, the difference between the two forecasts was as large as $8 - 10^0$ K for the 2 m temperature and $4 - 6$ g/kg for the 2 m moisture. Surface verification of T2 favors the MYJ run and shows that the YSU forecast is way too warm, while verification of Q2 favors the YSU forecast which has a reduced wet bias as compared to the

MYJ run. As a result, the two forecasts have errors in the 2 m relative humidity of similar magnitudes but opposite signs (too high in MYJ and too low in YSU). The YSU forecast exhibits an improved 10 m wind speed forecast.

For the case of 19 July 2010 similar differences exist between the two forecasts for the 2 m temperature and moisture. The MYJ forecast verifies better for 2 m temperature and YSU verifies better for 2 m moisture. For the 10 m wind speed, the MYJ forecast is better for the nighttime and the YSU forecast seems better during the daytime.

The diagnosed PBL heights in the MYJ and YSU forecasts are very different. At 20 UTC on 17 June 2010 the PBL height differences between the two forecasts were ~1.5 km in some regions (Fig. 3.13), with YSU predicting higher PBLs over most of CONUS. The PBL height in the YSU forecast is above 3 km in the Southwest U.S., obviously higher than commonly accepted. On the other hand, there are large regional variations over the CONUS in terms of the PBL height difference between the MYJ and YSU forecasts. We have also tested two modifications to the standard YSU PBL proposed by Yubao Liu and Greg Thompson of NCAR (personal exchanges). The modifications were found to mitigate the problem in YSU to some degree.

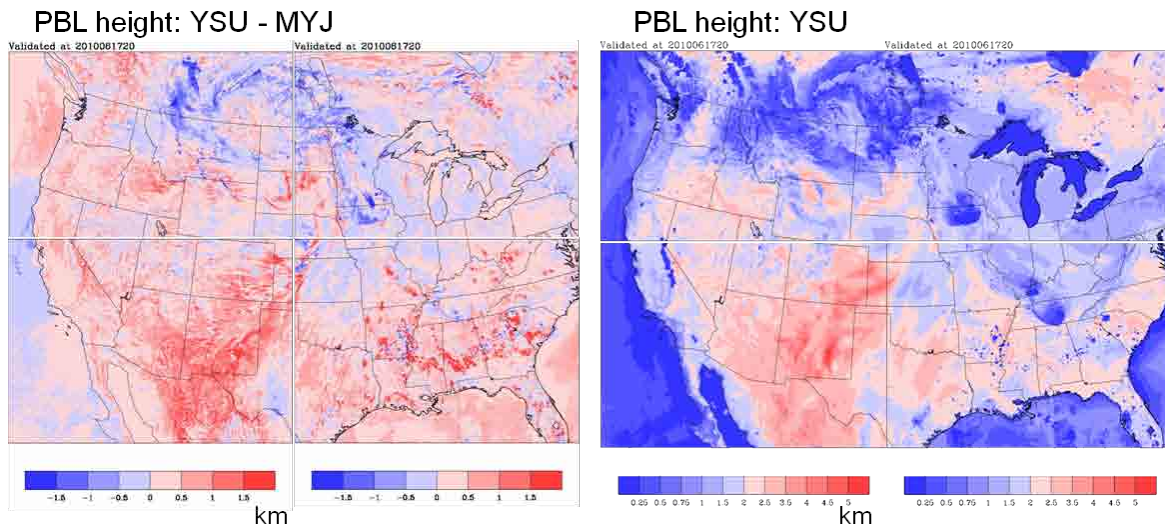


Figure 3.13: Left: differences in PBL height between the YSU and MYJ runs. Right: PBL height in the YSU run. The validation time is 20 UTC on 17 June 2010.

3.2.2 Forecast Sensitivity to the Microphysics Scheme

The Thompson microphysics scheme released with WRF v3.3.0 (hereafter referred as Thompson v3.3.0) is used in the baseline simulations. To find out the forecast sensitivity to microphysics schemes, three alternative microphysics schemes, namely the Milbrandt-Yau scheme, WSM6 scheme, and an updated version of the Thompson scheme (scheduled to be released with WRF v3.3.1 in September 2011, hereafter referred as Thompson v3.3.1), have been tested in 3 km sensitivity experiments. The experiments were performed for three cases, namely the 19 July, 16 September and 25 October 2010.

The Thompson v3.3.0 and v3.3.1 schemes (Thompson et al. 2008) are double-moment for rain

and single moment for ice, snow and graupel. The Milbrandt-Yau scheme (Milbrandt and Yau 2005) is a 7-class scheme that includes separate categories for hail and graupel with double-moment cloud, rain, ice, snow, graupel and hail. WSM6 stands for WRF Single-Moment 6-class scheme (Hong and Lim 2006), which has ice, snow and graupel processes. All four schemes are considered suitable for high-resolution simulations. It is found that the difference between Thompson v3.3.0 and Thompson v3.3.1 is significantly smaller than the differences between Thompson and Milbrandt-Yau or between Thompson and WSM6.

In general, the choice of microphysics scheme does not affect the “hit or miss” of the major convective systems in the forecast. The effect is more noticeable on the predicted structures of those systems, and may cause significant regional differences. Figure 3.14 shows that for verification hour of 00 UTC on 17 September 2010 the forecast using Milbrandt-Yau (correctly) predicts a stronger MCS near the Ohio-Pennsylvania border than the Thompson v3.3.0 scheme. The differences between the two model forecasts are evident. However the two forecasts are still closer to each other than they are to the observations.

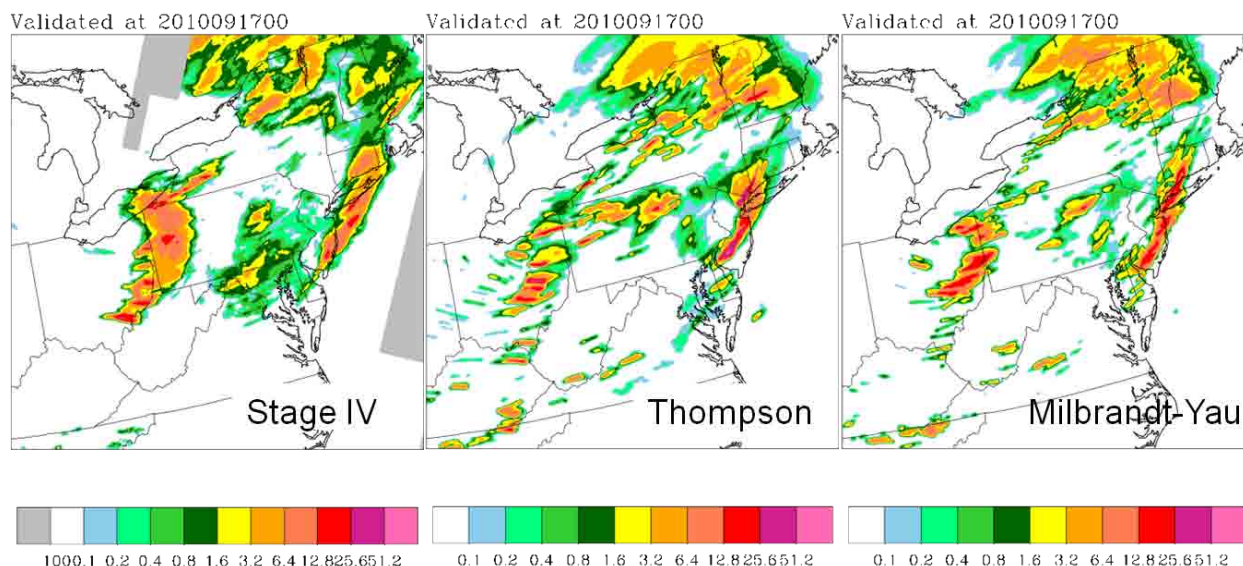


Figure 3.14: One-hour rainfall from Stage IV (left), and 3 km simulations using Thompson (center) and Milbrandt-Yau (right) microphysics schemes. The forecasts are initialized at 12 UTC on 16 September and valid at 00 UTC on 17 September 2010.

Among the microphysics schemes tested, none of these schemes seems obviously superior to the others in terms of rainfall verification scores (not shown). Rankings of the forecasts using the various schemes appear to be situation dependent. The Thompson schemes usually verify slightly better than the others for light precipitation (0.1 mm/h threshold), a result due to the fact that they usually produce more light precipitation. For heavier precipitation, the results vary case-to-case and depend on the verification hour.

More definite differences are found in the partitioning of microphysical species in the forecasts. Vertical profiles of the domain averaged graupel, snow and rain amount (e.g., Fig. 3.15) shows that the Milbrandt-Yau and WSM6 schemes produce much more graupels, in the air and

sometimes reaching the ground, than the Thompson schemes. The Thompson schemes produce more snow. The Milbrandt-Yau scheme also produces less ground precipitation than the other microphysics schemes in the case studies. An inspection of the rainfall maps shows that the Milbrandt-Yau scheme produced less areas of light precipitation and resulted in smaller values of domain-total rainfall.

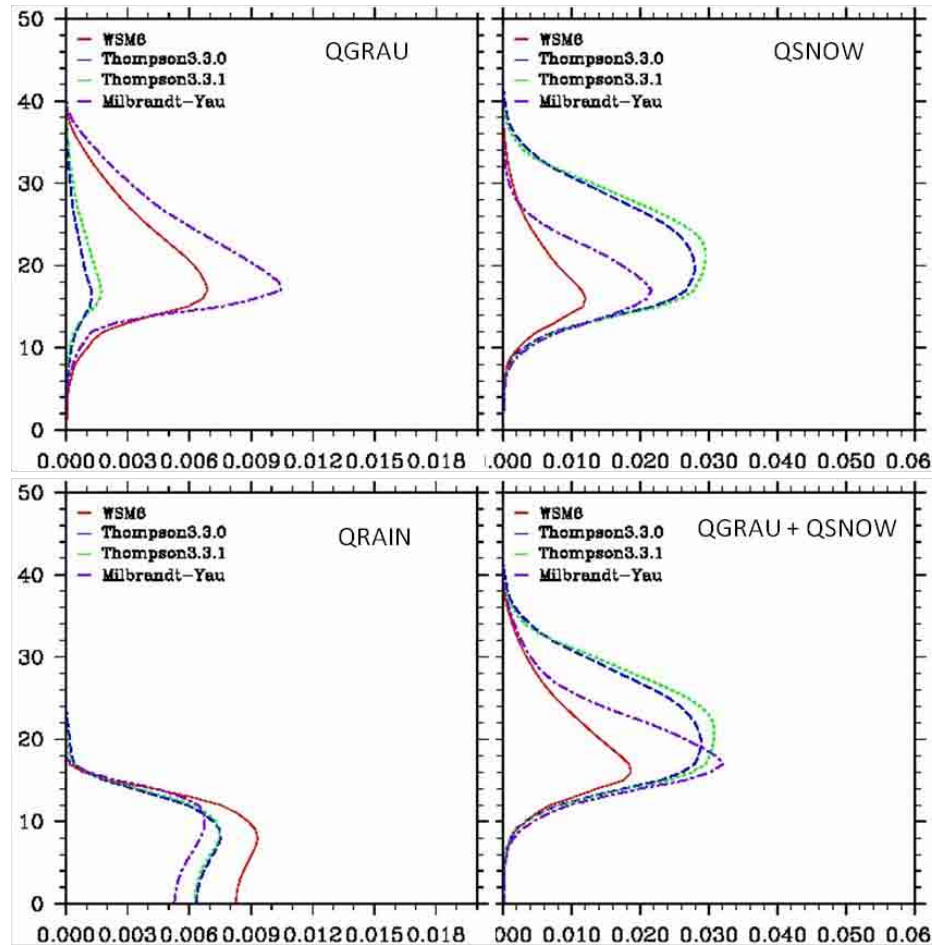


Figure 3.15: Vertical profiles of horizontally averaged microphysics variables (graupel, snow, rain and graupel+snow) in model runs with Thompson v3.3.0 (blue), Thompson v3.3.1 (green), Milbrandt-Yau (purple) and WSM6 (red) schemes. Shown are 12-hour forecasts valid at 0 UTC on 17 September 2010.

Surface and upper air verification shows that the Milbrandt-Yau scheme results in colder lower-level (below 700 mb) temperatures, leading to a larger cold bias in surface fields in the daytime forecasts. Meanwhile the 10 m wind speed forecast by the Milbrandt-Yau run is slightly better than the baseline run. The forecast sensitivity to the microphysics scheme is also seen in the simulated cold pools. An example is given in Fig. 3.16 of the cool pool behind the outflow boundary of the 16 September 2010 case. Comparing Fig. 3.16 with Fig. 3.9, it is found that the cold pool simulated in the Milbrandt-Yau is larger and stronger, though does not necessarily

verify better. A more detailed study of the microphysics schemes is beyond the scope of this report, but may be done later.

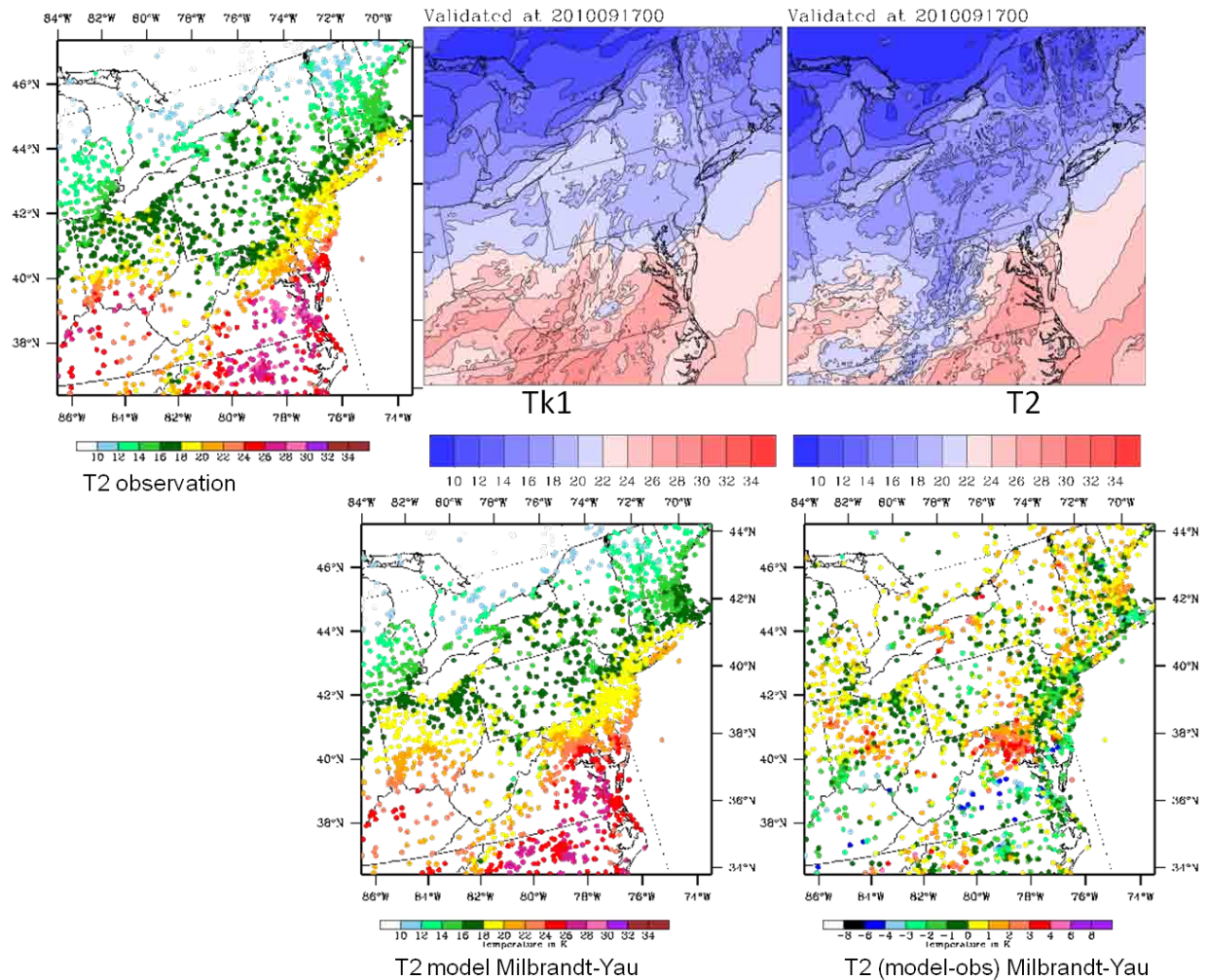


Figure 3.16: Observed 2 m temperature (T2 observation), modeled lowest-level (~ 8 m) temperature (Tk1), model derived 2 m temperature field (T2), model 2 m temperature at observation sites (T2 model Milbrandt-Yau), and discrepancies between the model and observed 2 m temperature (T2 model – observation Milbrandt-Yau), valid at 00 UTC on 17 September 2010 (12 hour forecast time). The forecast uses the Milbrandt-Yau microphysics scheme. Refer to Fig. 3.9 for plots of the baseline simulation.

3.2.3 Forecast Sensitivity to Land-Surface Model

For the case of 25 October 2010, a sensitivity experiment was conducted using the Noah Land Surface Model (LSM) to replace the RUC LSM in the baseline configurations. The choice does not seem to affect convection forecasts significantly. However, it causes some notable differences in the surface fields (e.g. the sensible and latent heat fluxes).

3.2.4 Forecast Sensitivity to Model vertical Resolution

A 13 km simulation with double vertical resolution was conducted for the case of 16 November 2010. In the baseline simulation, 50 vertical levels were used. In the sensitivity simulation, 100 vertical levels were applied. The resulted forecast of convection showed little differences (not shown). More will be discussed about the significance of doubling vertical resolution to the prediction of turbulence in Chapter 4.

3.3 **Assessment of real-time HRRR Forecasts of MCS during June – August 2010**

3.3.1 Data and Methodology

This section documents a study of the real-time High-Resolution Rapid Refresh (HRRR) performance on forecasting MCS during the summer of 2010, separate from the study on the seven joint cases. The data used for this study was provided by the Massachusetts Institute of Technology – Lincoln Labs (MIT/LL) and the HRRR model. The field used to identify Mesoscale Convective Systems (MCSs) was the vertically-integrated liquid water (VIL). The VIL field was visually inspected, frame-by-frame, with 15-min intervals for the period 1 June to 31 August 2010 (JJA). To make the study as objective as possible, we employed the following rules for determining convective initiation (CI) of a new MCS:

- 1) The CI must develop in an area where no ongoing convection was present within a radius of 300 km, with the exception of single isolated cells.
- 2) The area must have been free of convection within the past two hours.
- 3) The MCS must reach a maximum dimension of at least 100 km, as defined by the 3.5 kg m^{-3} VIL field, with no gaps greater than 10 km.
- 4) The CI time was defined as the time the MCS reached 100 km in length.
- 5) CI was limited to systems observed by the WSR-88D network of radars over the 48 contiguous United States.

These rules ensured that the cases of MCS CI only accounted for clear-cut initiation, and didn't include regions of expanding or intensifying pre-existing convection.

Nearly 70% of the 93 identified occurrences of MCS CI during JJA 2010 occurred between the hours of 12 and 18 local solar time (LST), while nearly 20% occurred between the hours of 18 and midnight (Fig. 3.17). LST was estimated by the following equation:

$$\text{LST} = \text{UTC} + \text{Lon}_{\text{CI}} \times (15^\circ \text{ hr}^{-1})^{-1}$$

where UTC is universal coordinated time and Lon_{CI} is the longitude of the center of the MCS at the time of CI. These CI events were located most often over the central and northern Great Plains (Fig. 3.18), with lesser counts over the South and East.

Four different issue times of the HRRR (00, 06, 12 and 18 UTC) were used to determine the model performance in producing MCS CI. Forecast hours 3 to 15 of the 15-hour HRRR simulations were used. The first three forecast hours were not used because the model was in the process of spinning up. The performance of the HRRR in creating MCS CI was categorized into five categories, as follows:

- 1) Hit: The distance between the centroid of the observed MCS and the simulated MCS at the time of CI was within 100 km and the time difference of CI was one hour or less.
- 2) Hit within 1 to 2 hours: The same as the “Hit” classification, except that the timing of CI was off by more than one hour and less than or equal to two hours.
- 3) Partial Hit: The distance between the centroid of the observed MCS and the simulated MCS at the time of CI was between 100 and 300 km and the time difference of CI was greater than two hours, but less than five. In addition, widespread convection not fitting our definition of MCS was also included as a partial hit.
- 4) Miss: The HRRR simulation failed to produce at least a partial hit within 300 km and/or within five hours.
- 5) False Alarm: The HRRR simulation produced an MCS, but the observed VIL did not show a MCS within 300 km and/or within five hours.

The number of identified MCSs within the 00, 06, 12 and 18 UTC HRRR simulations for JJA 2010 are 12, 31, 85 and 56, respectively.

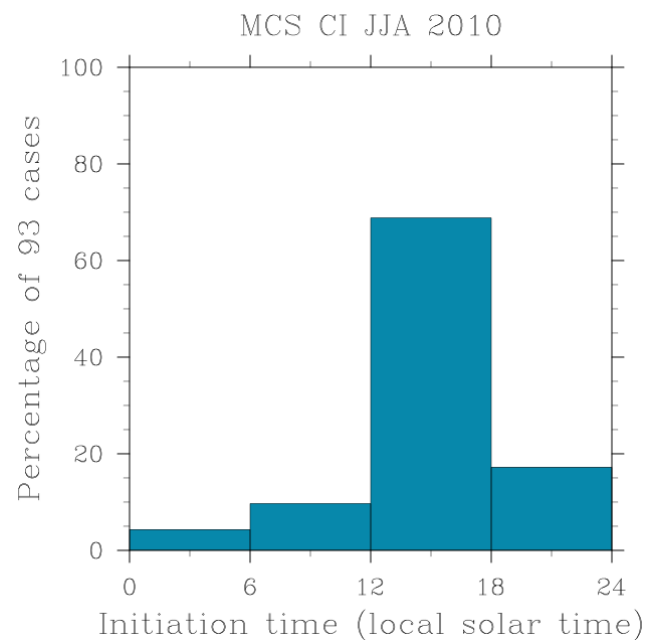


Figure 3.17: Observed MCS convective initiation by time range.

In order to substantiate the results from the more subjective analysis just described, we are also in the process of using an automated system to identify new MCS CI. The rules within the automated processing were made to mirror as closely as possible those from the subjective analysis. The following steps are being followed in order to identify the observed cases of MCS CI:

- 1) In the areas where ongoing convection was present, a mask was applied to the extrapolated VIL fields for 2 hours into the future.

- 2) All observed VIL within this masked area was set to missing.
- 3) The remaining VIL field was expanded by 5 km in all directions, to account for gaps up to 10 km between storms in an MCS.
- 4) A storm-tracking program called Titan was used to identify new MCS CI greater than 100 km in length.

MCS CI locations, colored by time of initiation (local solar time)

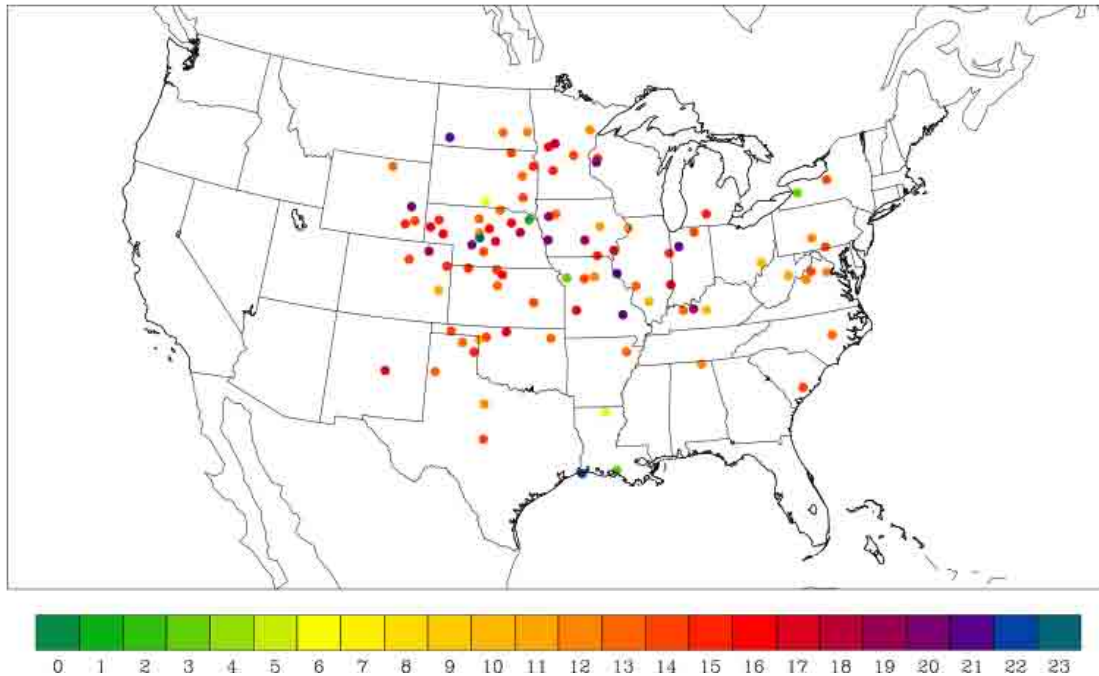


Figure 3.18: Locations of MCS convective initiation during JJA 2010, colored by time of initiation (UTC).

3.3.2 Results

The HRRR simulations were assessed to determine their skill in the prediction of MCS CI using the five categories discussed in the previous section. The total number of cases, including all types of hits, misses and false alarms were 18, 41, 94 and 60, respectively for the 00, 06, 12 and 18 UTC HRRR simulations. Of these cases, one-third was classified as hits, being within 100 km and 1 hour of the observed MCS CI location and time (Fig. 3.19). Extending the time window to 2 hours, half of the HRRR simulations produced a hit, while including all partial hits resulted in a 74% success rate. Only 12% and 14% were pure false alarms and pure misses, respectively. When considering the HRRR performance by issue hour, it is seen that the 12 and 18 UTC HRRR simulations performed the best, getting at least a partial hit 78% of the time. The overnight 00 and 06 UTC simulations, with notably smaller sample sizes, fared slightly worse with only a 67% partial hit success rate at best.

HRRR Model Performance MCS Initiations - JJA 2010

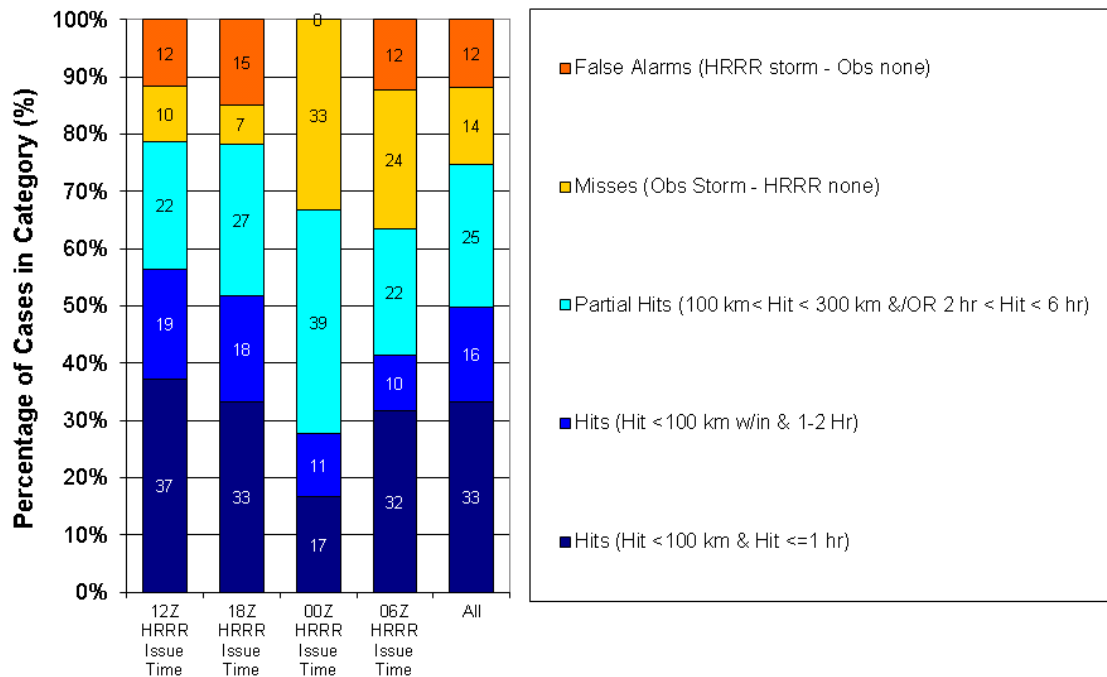


Figure 3.19: Performance of HRRR simulations by model issue time.

When considering the HRRR performance by lead time, there were 41, 51, 41 and 40 cases in the 3 – 6, 6 – 9, 9 – 12 and 12 – 15 forecast hour time windows. The most notable overall trend in the skill with forecast lead time was the increase of pure misses with each increasing forecast time window (Fig. 3.20). As for the timing of CI between the HRRR MCSs and those observed, there exists a near-Gaussian distribution for all simulations, with the exception of the 00 UTC HRRR, which has a much more random distribution (Fig. 3.21). This is further seen in calculations of the kurtosis, which for the 06, 12 and 18 UTC cases are 3.2, 2.1 and 2.1, respectively, but is -1.1 for the 00 UTC case. There are also variations in the timing with respect to the LST (Fig. 3.22). For the afternoon cases (12 – 18 LST), the mean difference between the observed and HRRR-simulated MCS CI was 1.3 hours; whereas during the first half of the night (18 – 00 LST), the mean difference was over three times as much (4.5 hours). This is likely the result of the better-resolved surface forcing of the daytime MCSs and less-resolved forcing aloft of the nocturnal MCSs. The objective portion of the study is still ongoing and will be evaluated in the near future.

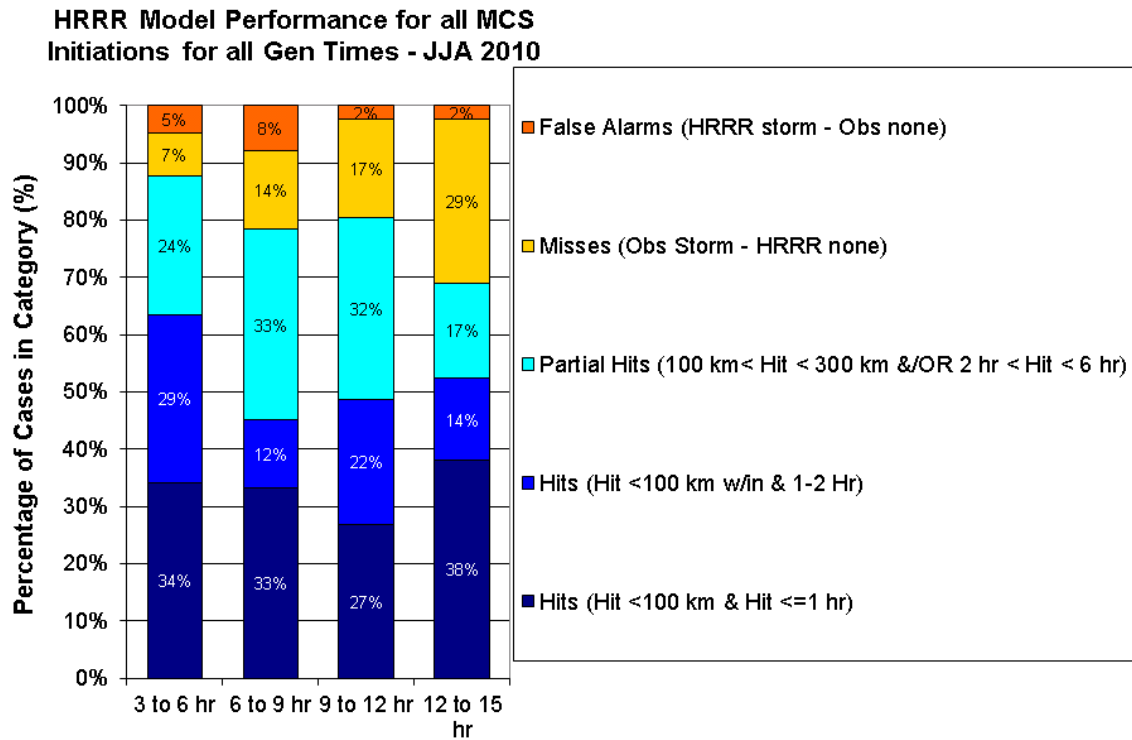


Figure 3.20: Performance of HRRR simulations by lead time.

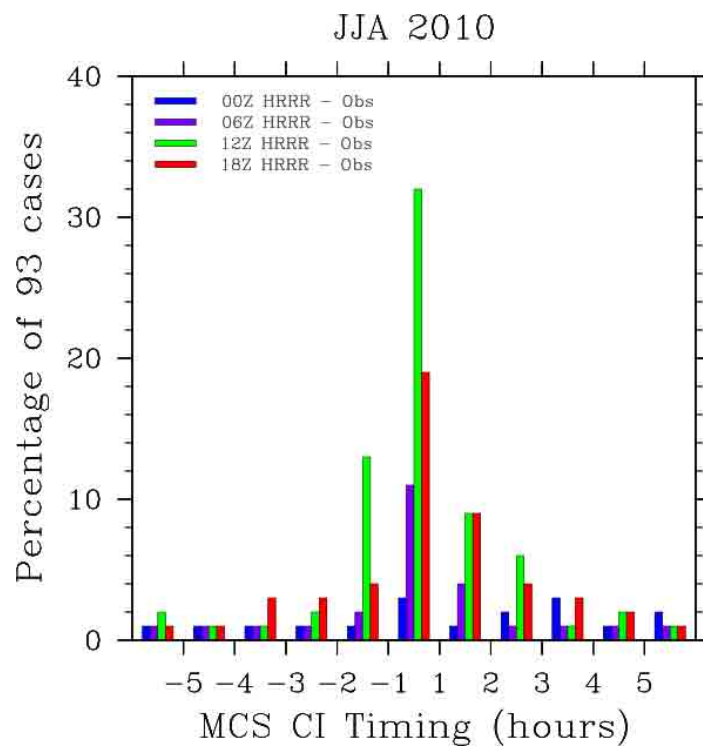


Figure 3.21: The time difference between the HRRR and observed MCS CI, binned by time difference.

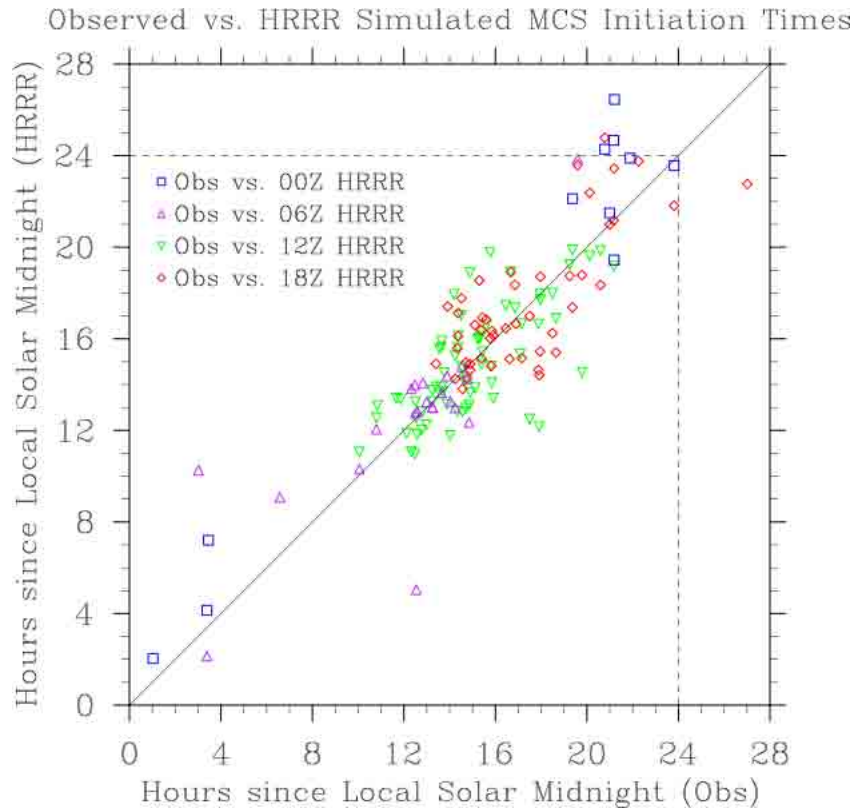


Figure 3.22: Scatter plot of HRRR MCS CI times versus observed MCS CI times.

3.4 Summary of convective Analyses and Suggestions for future Work

Riding on the recent developments in WRF and high resolution data assimilation techniques, the 3 km HRRR forecasts did reasonably well on convective systems of relatively large spatial and temporal extent, such as lines of convection, regional scattered convection, and winter storms. It usually places the systems in roughly the same region as observed, depicts storm structures with realism, and predicts the convective mode correctly. On the other hand, errors are evident in the predicted mesoscale details, such as in the exact location and orientation of line storms and storm clusters, initiation and maintenance of MCSs, and timing and strength of sea breeze convection. Such errors can be significant even in very short-term, say 1 – 3 hour forecasts.

A problem of over-prediction was identified with the 13 km simulations. The model also seems to experience a “spin-down” period in the first hour. Large areas of false-alarm precipitation are predicted immediately following the initialization, especially over the ocean in the southeast part of the domain. An explanation may be that since the RUC tends to under-forecast convection, the RUC analyses that are tuned up for RUC may be too hot for starting WRF forecasts.

Quantitative skills of the 3 km baseline simulations show cold and wet biases in surface temperature and moisture, and over-forecasts of surface wind speed. The errors in temperature and moisture gradually grow with forecast length and correlated with the diurnal cycle. The wind speed errors grow rapidly in the first forecast hour, and the wind direction errors (RMSE) are the largest in the initial conditions. This may be an indication of small-scale mismatches and

imbalances in the initial conditions. Surface observation also reveal that the simulated temperature gradients across sharp boundaries tend to be smoothed. Verification using observed soundings and surface observations revealed obvious model deficiencies at low levels in terms of wind direction, temperature, and humidity. Detailed studies on how to improve the analysis and forecast of low-level representations in models are key to improve short-term convective forecasting.

Sensitivity experiments show that, using different physics options usually does not adversely change the forecast of convection. However, prediction of the details of storms can be significantly impacted. Moderate sensitivity is seen in rainfall forecasts to PBL and microphysics schemes. On the other hand, it is not clear which scheme has better skills for convective forecasting over the entire CONUS. The skills of the physics packages may be regime dependent. It is also found that the physics option may affect certain model fields (e.g., microphysics scheme affects hydrometeor partitions) more than they do to rainfall. The YSU PBL scheme is found to severely over-estimate PBL height and therefore PBL mixing. Two modifications to the standard YSU were therefore proposed (Yubao Liu and Greg Thompson of NCAR, personal communications) and were found to mitigate the problems to some degree.

Post-season analysis found that the operational HRRR forecasts missed more than 30% of the nightly MCS initiation during June-August 2010. To better understand and grasp the forcing of nocturnal MCSs and therefore to improve the forecast will continue to be a challenge for the HRRR model.

For the next year, we plan to further analyze the baseline and sensitivity simulations performed in FY11. Attention will be given to the evaluation of simulations focused on regional domains and individual convective systems. Additional sensitivity experiments will be conducted with regard to the model initial conditions and physics schemes. These experiments and analyses will also focus on a particular forecast challenge, such as the initiation and up-scale growth of MCS on 19 July 2010. Previous experiments have shown that model initial conditions are still the controlling factor in convective forecast skills. The verifications against observations in this study provided further evidence that the current model has significant problems with the boundary layer, especially for wind and temperature; both are sensitive fields for convective initiation. We believe that the key for improving the analyses and, therefore, the forecasts of these fields and rainfall, are to improve the surface data and radar data assimilation in concert with improving the boundary layer physics.

References for Section 3

- Hong, S-Y., and J-O. J. Lim, 2006: The WRF single-moment microphysics scheme (WSM6). *J. Korean Meteor. Soc.*, **42**, 129 – 151.
- Hong, S-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318 – 2341.
- Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous layer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927 – 945.
- Janjic, Z. I., 2001: *Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP Meso model*. NOAA/NWS/NCEP Office Note 437, 61 pp.

- Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV hourly precipitation analyses: development and applications. *Preprints, 19th Conf. on Hydrology*, American Meteorological Society, San Diego, CA, 9-13 January 2005.
- Milbrandt, J. A., and M.-K. Yau, 2005: A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *J. Atmos. Sci.*, **62**, 3051 – 3064.
- Roberts, N. M., 2005: *An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall*. Met Office Tech. Rep. 455, 80 pp.
- Sukoriansky, S., B. Galperin, V. Perov, 2005: Application of a new spectral theory of stable stratified turbulence to the atmospheric boundary layer over sea ice. *Boundary-Layer Meteorol.*, **117**, 231 – 257.
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. *Mon. Wea. Rev.*, **136**, 5095 – 5115.

4. Analyses focused on Turbulence

The affect of grid resolution on turbulence diagnostics was evaluated using a combination of case study and statistical analyses. In addition some sensitivity studies to model configuration were executed. Only a limited number of cases (8) were available that contained output data from both the 13 km and 3 km resolution grids, but this appears to be sufficient to draw at least qualitative conclusions about the affects of resolution.

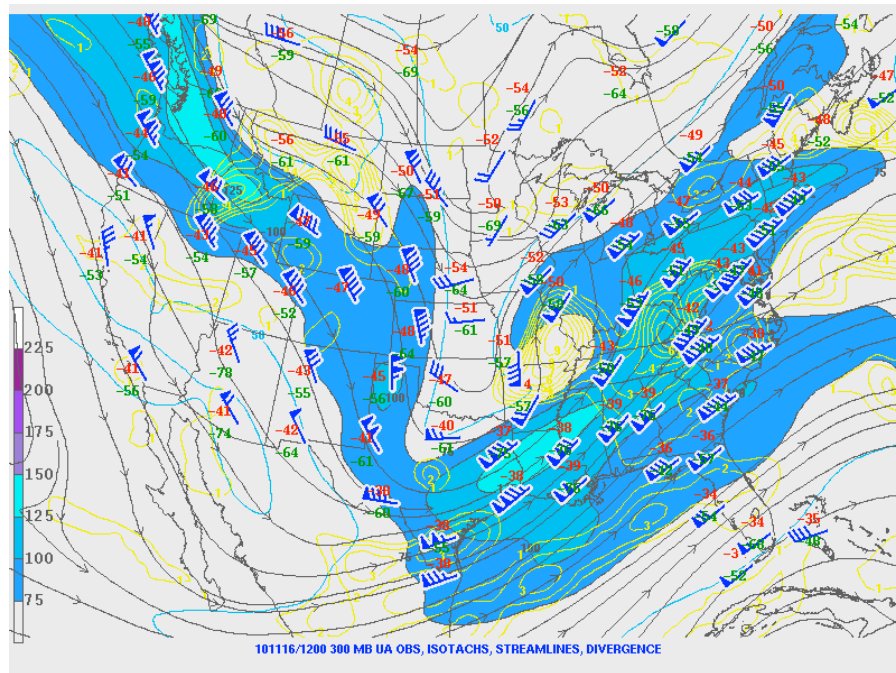


Figure 4.1: Surface analysis/satellite composite for 16 November 2010 at 1330 UTC (upper frame) and 300 hPa wind analysis valid at 1200 UTC (lower frame).

4.1 Qualitative Assessment of Baseline and Sensitivity Runs

Differences between the 13 km and 3 km model output turbulence diagnostic fields were analyzed in detail for one particular case: 16 November 2010. The surface/satellite composite and 300 hPa analysis at midday for this case is shown in Fig. 4.1. Note the presence of a frontal system moving eastward over the SE, associated with strong convection throughout most of the troposphere and a strong jet aloft. These features contribute to several severe PIREPs during the day as shown in Fig. 4.2. Also of note in Fig. 4.2 are dense PIREPS due to mountain wave turbulence (MWT) over the Colorado/New Mexico Rocky Mountain region and the Wasatch Range. Thus, this case contains sources of turbulence due to convection, upper level clear-air turbulence (CAT) and mountain wave turbulence (MWT). Both 6-hour and 12-hour forecasts were evaluated for both the 3 km and 13 km grids. In addition, sensitivity studies were performed by a) increasing the number of vertical levels used from 50 to 100, b) changing the boundary layer parameterization scheme from the operational MYJ (Mellor-Yamada-Janjić, Mellor and Yamada 1982, Janjić 1990) to the YSU scheme (Yonsei University, Hong et al.

2006), and 3) reducing the 6th-order numerical diffusion coefficient to 0 from its baseline value of 0.25.

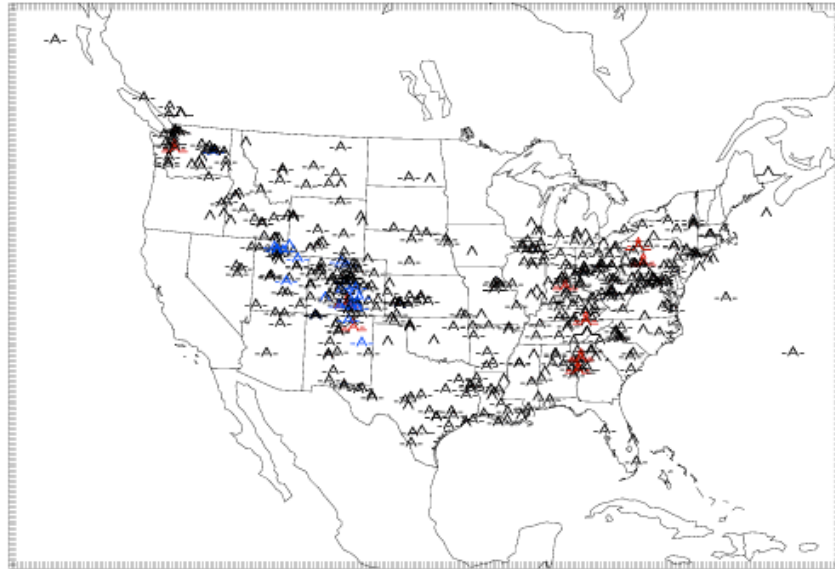


Figure 4.2: PIREPs for 16 November 2010 during the 24-hr time period from 0 UTC – 2400 UTC through the layer 10,000 ft – 45,000 ft. Red indicates severe reports, blue indicates mountain wave reports.

Figures 4.3 – 4.7 provide comparison plots of selected turbulence diagnostic fields at FL330 for 6-hour forecasts from the sensitivity runs. A GTG-combination (Sharman et al. 2006) based on weights and a suite of diagnostics used with the WRF-RR is included as Fig. 4.6. For the 13 km runs the results are insensitive to the model configuration, at least for the configurations tested. Note that increasing the number of vertical levels did not change qualitatively the patterns, but it did produce higher values of the diagnostics overall. Note that in general the MYJ turbulence kinetic energy (TKE) under-predicts the turbulence levels, but a change in threshold may help this. Also the 3 km pattern is much more structured than the 13 km pattern, but overall, qualitatively it looks the same as the 13 km output.

To get an idea of the affect of the model configuration on the vertical structure, Fig. 4.8 provides a series of Ri profiles derived from 12-hour forecast fields using the nearest grid point to Denver (DEN, taken as 39.75N, 104.87W). The DEN sounding taken at the valid time (0Z 17 Nov 2010) is shown for comparison. Of significance for turbulence is the presence of low regions of Ri. The sounding indicates regions of reduced Ri occur in layers from about 5 – 9 km and 12 – 15 km elevation. These two layers are reproduced in all the simulation variants, but in general they are too high and not as deep, and this is true even for the 100 level (13 km) and the 3 km runs. All variants are also substantially smoother than the sounding profile, so that many of the individual peaks in the observed profile are missed.

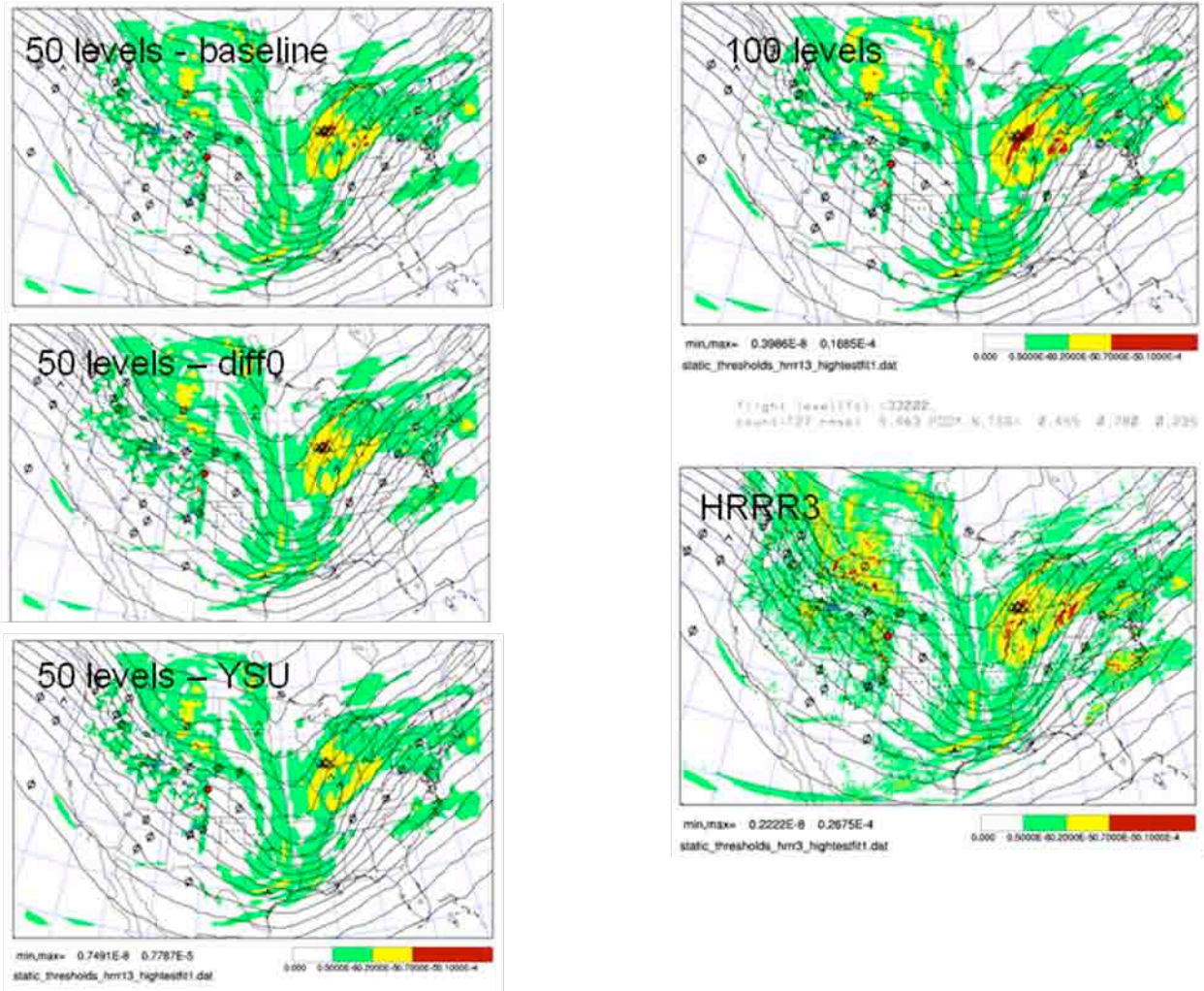


Figure 4.3: Ellrod1 turbulence index at FL330 from 5 different model configurations: 13 km baseline WRF-RR emulation, 6th order diffusion = 0 (13 km), YSU PBL scheme (13 km), 101 vertical levels (13 km), and 3 km baseline HRRR emulation. Contour intervals are the same on each plot, and are chosen such that no color indicates expected smooth, green indicates light turbulence, orange indicates moderate turbulence, and red indicates regions of expected severe turbulence. PIREPS and in-situ data near the valid time are also displayed in the standard symbology.

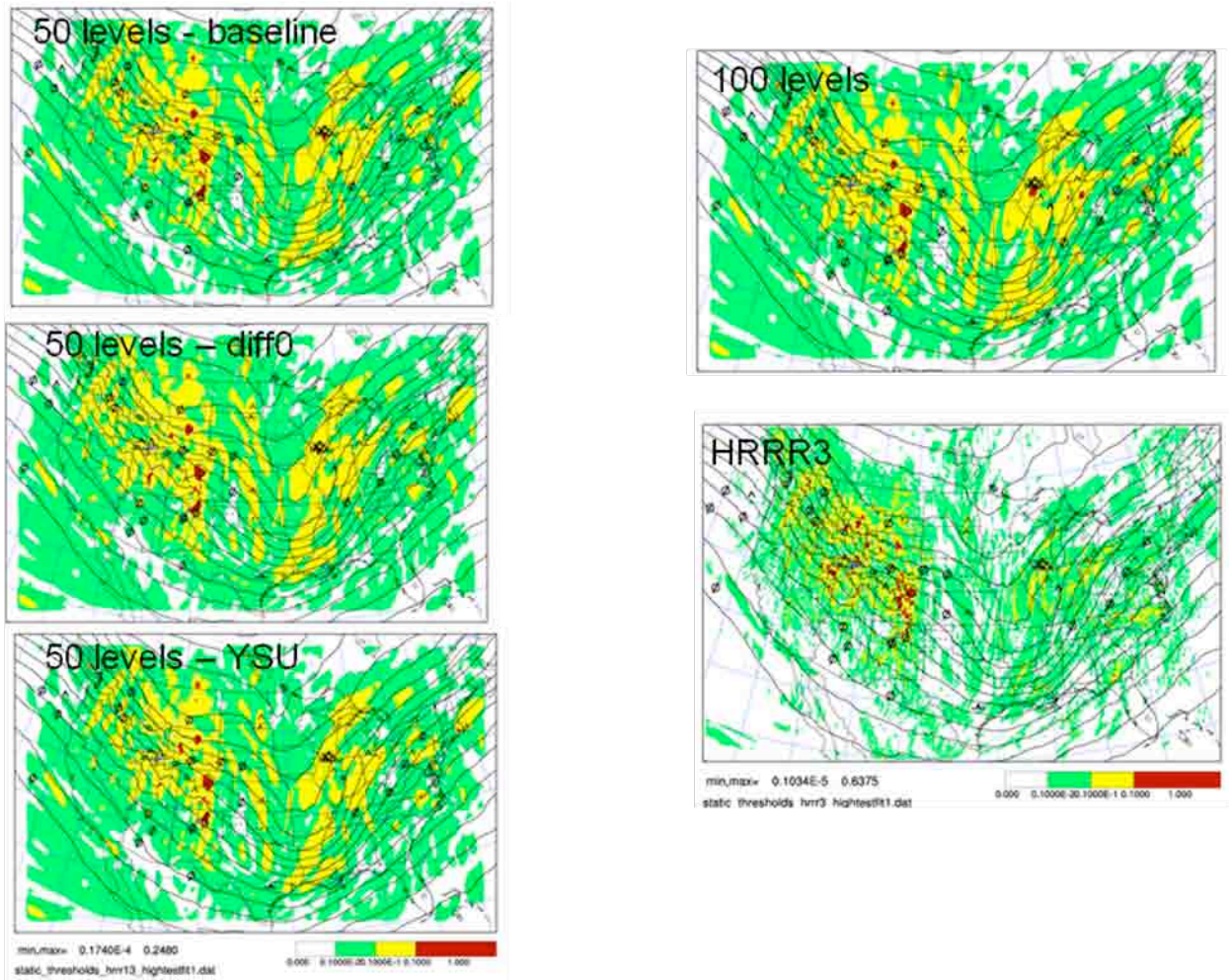


Figure 4.4: Same as Fig. 4.3 except that the EDR/Ri turbulence index is displayed.

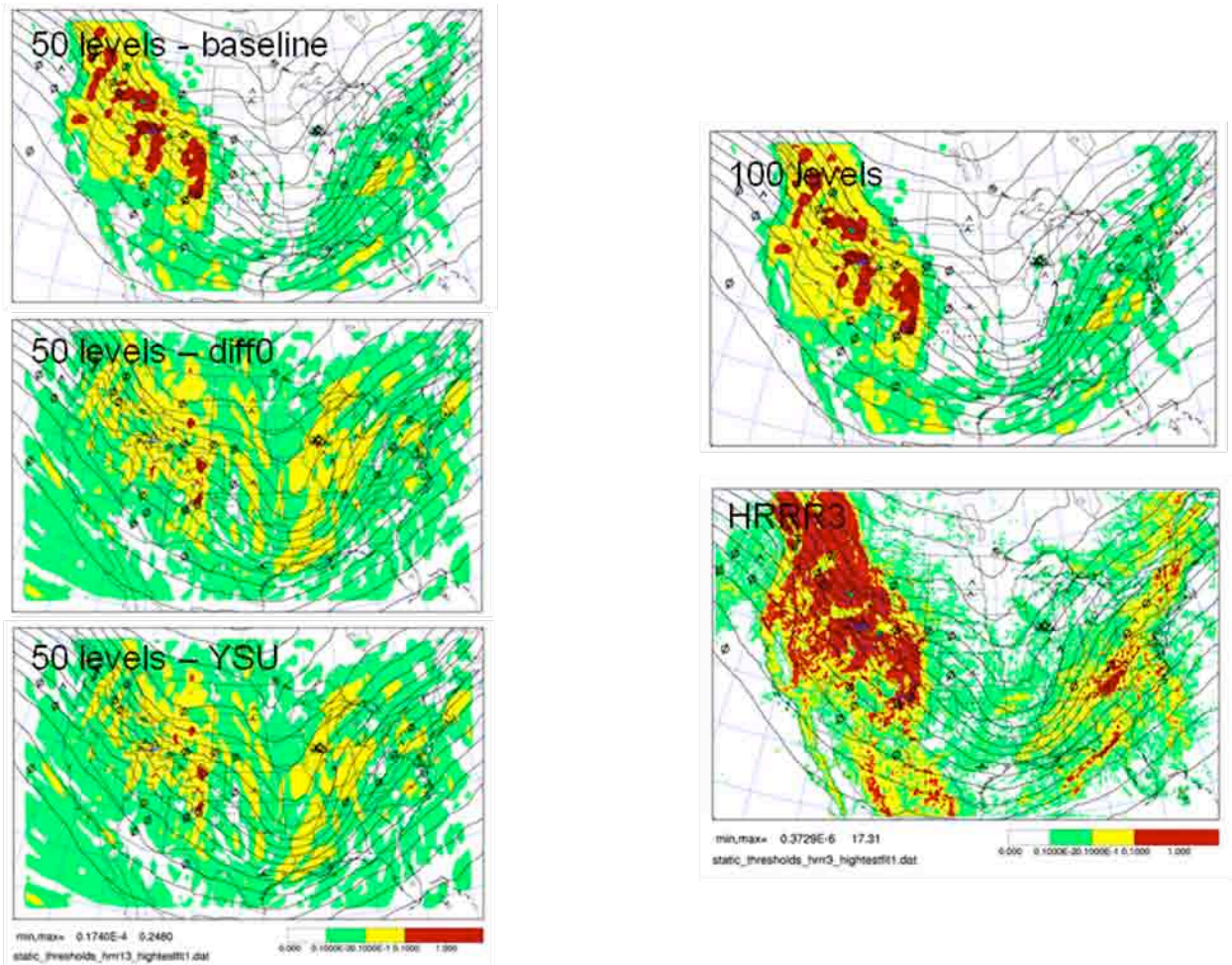


Figure 4.5: Same as Fig. 4.3 except that the SIGW/Ri turbulence index is displayed.

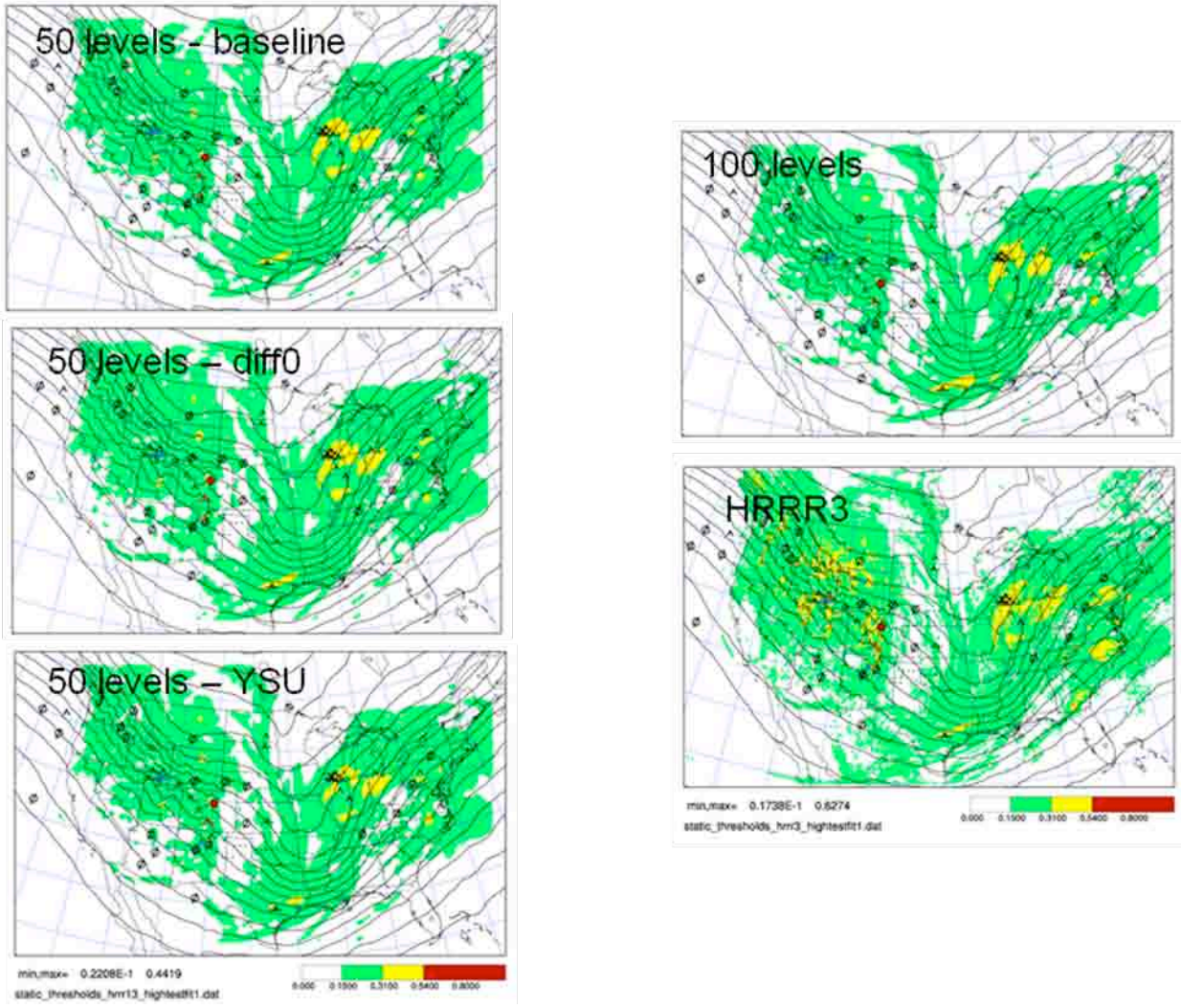


Figure 4.6: Same as Fig. 4.3 except that GTG combination is displayed.

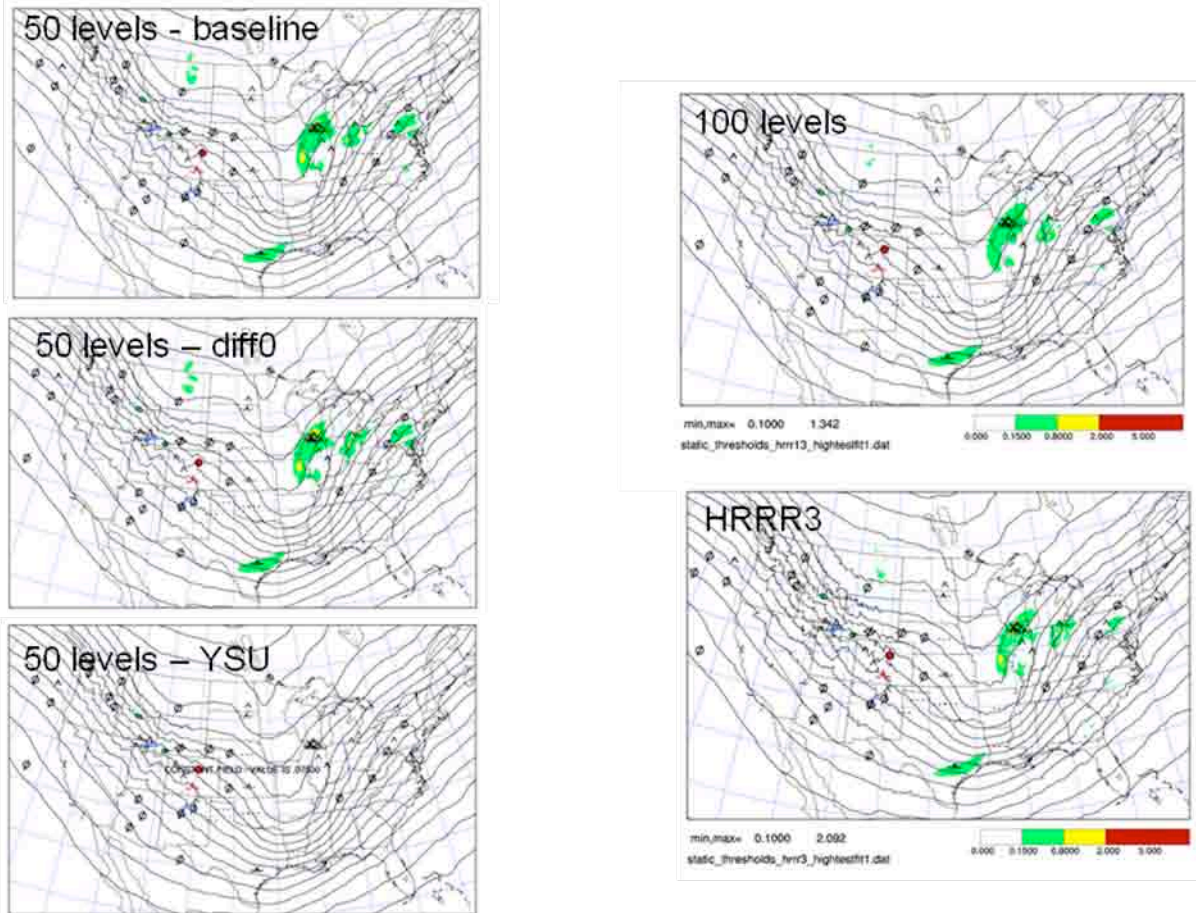


Figure 4.7: Same as Fig. 4.3 except that the MYJ subgrid TKE is displayed.

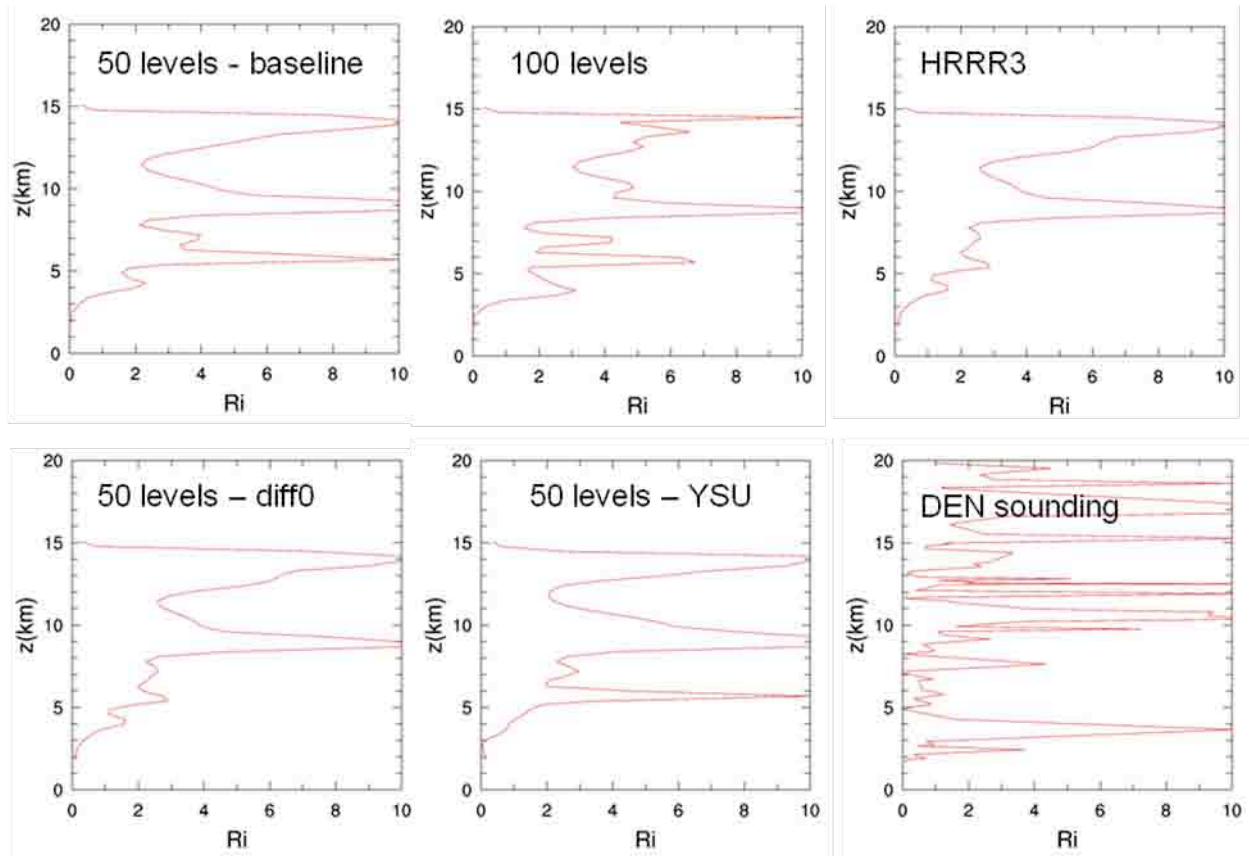


Figure 4.8: Vertical profiles of Ri from 12-hour forecasts at the closest grid point to DEN (39.75N, 104.87W) for various model configurations compared to the DEN sounding valid at the same time (0 UTC 17 November 2010).

4.2 Quantitative Assessment of Baseline Runs

Performance statistics were derived from 6-hour and 12-hour forecast of the 13 km and 3 km baseline configurations. A total of eight forecasts were computed for the following days:

2009-10-09
 2009-12-09
 2010-02-04
 2010-06-17
 2010-09-16
 2010-10-25
 2010-11-16
 2011-02-01

Table 4.1: Comparison of AUCs for selected turbulence diagnostics, the GTG combination, and the MYJ subgrid scale turbulence parameterization scheme.

Index	AUC			
	WRFRR13	HRRR3 4 pt avg.	HRRR3 12 pt avg.	HRRR3 20 pt avg.
	6-hour forecast			
Ellrod1	0.8149	0.8075	0.8093	0.8107
Brown2	0.8162	0.8194	0.8202	0.8209
Fth	0.8313	0.8196	0.8225	0.8252
EDR/Ri	0.8252	0.8037	0.8057	0.8073
MYJ SGS TKE	0.6756	0.6511	0.6605	0.6685
GTG	0.8507	0.8367	0.8382	0.8395
	12-hour forecast			
Ellrod1	0.7872	0.7436	0.7498	0.7557
Brown2	0.7686	0.7383	0.7426	0.7471
Fth	0.7985	0.7781	0.7801	0.7820
EDR/Ri	0.7857	0.7409	0.7458	0.7512
MYJ SGS TKE	0.6184	0.5984	0.6026	0.6175
GTG	0.8019	0.7663	0.7686	0.7716

For each day the 6-hour and 12-hour turbulence diagnostic forecast fields were compared to observations from PIREPs and in situ data within 1.25 hours of the valid time in the FL200 – FL450 layer. This provided 9476 observations for the 6-hour forecast (valid 1800 UTC) and 8673 observations for the 12-hour forecast (valid 0 UTC the next day) for comparison. With these observations 2x2 contingency tables were developed and PODs computed. For this exercise the PODs computed were for the probability of correctly predicting smooth turbulence conditions (PODN) and the probability of predicting moderate-or-greater (MOG) turbulence conditions (PODY). For different thresholds a curve can be constructed that traces out the PODY – PODN tradeoffs, called a ROC (Receiver Operating Characteristics) curve. A standard discrimination skill statistic is the area under this ROC curve (AUC), and the higher the AUC, the greater the overall discrimination skill (see e.g., Marzban 2004, Sharman et al. 2006). But the comparison of high-resolution grids to lower-resolution grids is problematic, with the lower-resolution grids usually providing higher forecast skills for the wrong reasons (e.g., Mass et al. 2002, Davis et al. 2006). To account for this affect, three averaging domains for the 3 km grids were used. The first compares the observed turbulence value to the average of the turbulence diagnostic values at 4 grid points (at the observation flight level) surrounding the observation. The second averaging domain involves comparing each observation to the 12 grid points

surrounding the observation, and the third uses 20 points surrounding the observation. This in effect provides an average over a $(3 \text{ km})^2$, $(9 \text{ km})^2$, and $(15 \text{ km})^2$ area for the three averaging options, respectively, compared to the $(13 \text{ km})^2$ average on the 13 km grid. Thus the averaging domain is closest for averaging option 3.

Table 4.1 shows the AUCs for both the 6-hour and 12-hour forecasts for selected turbulence diagnostics, the GTG combination, and the MYJ parameterized TKE. These results indicate that

- a) there is no clear advantage using the 3 km grid, even when the 20 point averaging of the 3 km grid values are used;
- b) averaging does help the results;
- c) the 6-hour forecast skills are superior to the 12-hour forecast skill, about 6%;
- d) the MYJ parameterized SGS TKE is not very skillful for either the 3 km or 13 km grids.

4.3 Summary of Turbulence Analyses and Suggestions for future Work

- Different model configurations have different affects on different diagnostics;
- Largest affect is with 3 km horizontal grid;
- Doubling vertical resolution everywhere makes little difference;
- Diff = 0 makes little difference;
- YSU PBL makes little difference except it provides no upper level SGSTKE;
- Limited statistical evaluations indicate the overall forecast skill is no better for the 3 km grid than for the 13 km grid;
- The MYJ turbulence parameterization performs very poorly compared to post-processed turbulence diagnostics for both the 3-km and 13-km grids.

It is not clear how to explain these results, especially the observation that higher resolutions, both horizontal and vertical, do not lead to better results, either in the case studies or the statistical analyses. This might be in part due to the cases chosen, which were primarily driven by convection considerations. Due to the highly transient nature of convectively-induced turbulence, these cases would be more difficult to handle from the turbulence perspective.

Perhaps future work should concentrate on cases that are more of the CAT variety. Also, this study did not consider mid- or upper-levels. It could be that higher resolution is more beneficial for turbulence sources active in this part of the atmosphere, but this also needs to be investigated. Finally, it is not clear why the MYJ scheme does not provide better statistical performance. This needs to be investigated further as well.

References for Section 4

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I. Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **112**, 918 – 922.

- Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318 – 2341.
- Janjic, Z. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.*, **118**, 1429 – 1443.
- Marzban, C., 2004: The ROC curve and the area under it as performance measures. *Wea. Forecasting*, **19**, 1106 – 1114.
- Mass, C. F., D. Ovens, K. Westrik, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? The results of two years of real-time numerical weather prediction over the Pacific Northwest. *Bull. Amer. Meteor. Soc.*, **83**, 407 – 430.
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. and Space Phys.*, **20**, 851-875.
- Sharman, R., C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268 – 287.

5. Analyses focused on In-flight Icing

In-flight icing occurs when supercooled liquid water drops impact an aircraft while in flight. Accurate numerical weather predictions of the presence and amount of supercooled liquid water (SLW) require accurate predictions of water vapor, vertical velocity and the resulting condensate. Knowledge of the locations and times when SLW is present are available primarily through pilot reports (PIREPs) that mention icing conditions. In addition to providing the location of the icing encounter, a PIREP includes the pilot's estimate as trace, light, moderate or severe. The accuracy of PIREP locations and times are of reasonable quality to use for evaluation and comparison of numerical weather prediction models and associated algorithms (Bernstein et al. 2005). In this section we test the 13-km (emulating the WRF-RR) and 3-km (emulating the HRRR) baseline WRF runs and their ability to predict icing conditions.

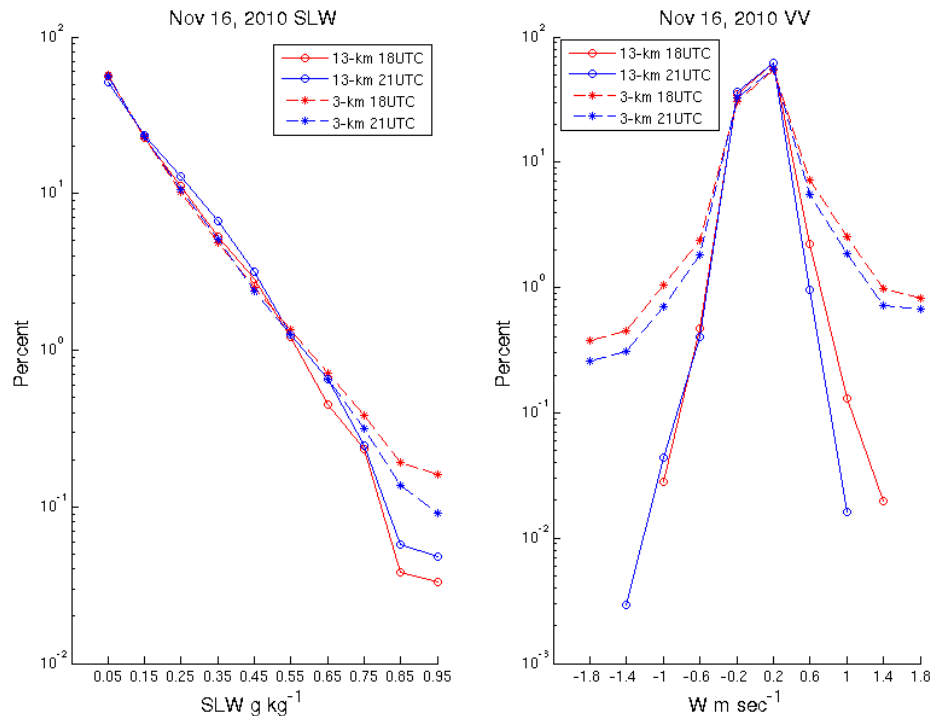


Figure 5.1: Percent distributions of supercooled liquid water (left) and co-located vertical velocity (right) at model locations where supercooled liquid water is predicted by the model. The 6 and 9-hour forecasts at both 3 and 13 km are shown for 16 November 2010.

5.1 Quantitative Analysis

In the initial quantitative analysis we compared the relative volume and amount, as well as the vertical velocities, of the model gridpoints with predicted supercooled liquid water (SLW). The 6 and 9-hour forecasts for the 3 and 13-km WRF baseline configurations were compared. In the winter cases the relative percent of grid points predicting SLW

was closely matched for both models until the final 0.5% of the distribution, where the 3-km predicted a higher percentage of gridpoints which contained a larger amount of SLW (Fig. 5.1). The range of vertical velocities (W) where SLW was predicted was quite similar in both models and forecast times. The majority of SLW predictions occurred when W has $\sim 0.2 \text{ m sec}^{-1}$ updrafts, although a significant number occurred in slight downdrafts. The 3-km predicted ascent and descent rates greater than 1 m sec^{-1} more often than the 13-km WRF. These stronger ascent and descent values occurred at less than 2% of the SLW gridpoints. These results were also similar in the other winter cases and are not shown. In the summer cases most (90%) of the SLW gridpoints in both models had $\text{SLW} < 0.3 \text{ g kg}^{-1}$ (Fig. 5.2). The 3-km model predicted SLW exceeding 0.3 g kg^{-1} much more often than the 13-km. The 9-hour forecast valid at 2100 UTC predicted greater amounts of SLW than at 1800 UTC in both the 3 and 13-km models. This may have been due to microphysical spin-up or more likely the stronger insolation present at 2100 UTC in the North American summer.

Similar analysis was also done across four sub-regions (NW, SW, NE, and SE) to test the role of the 3-km higher-resolution terrain in producing stronger vertical velocities and additional condensate. The results of this study were inconclusive for the cases studies selected and therefore are not shown.

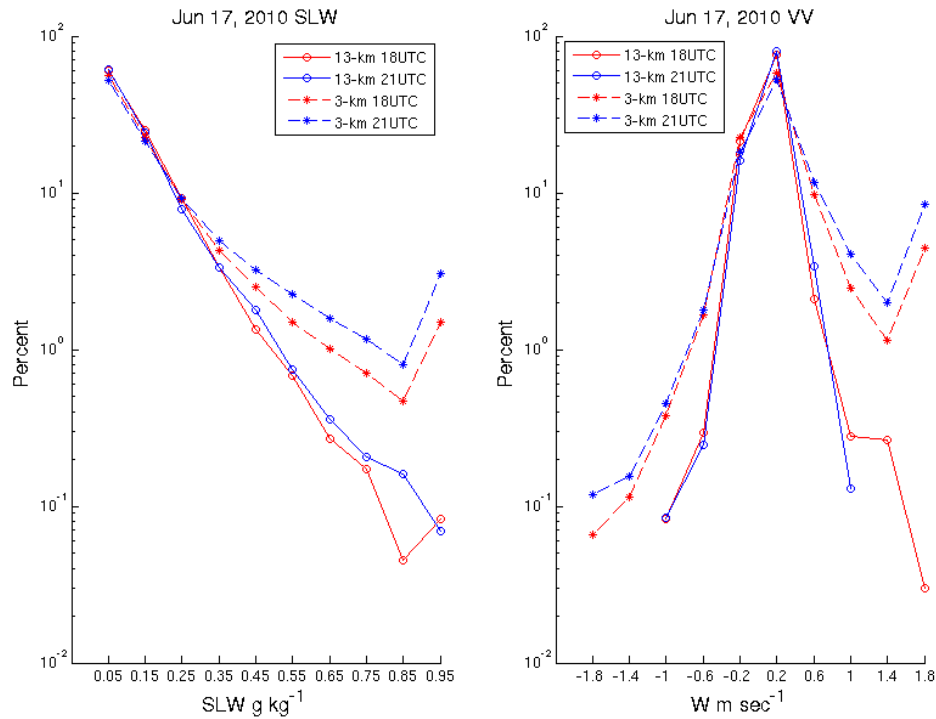


Figure 5.2: Distributions of supercooled liquid water (left) and co-located vertical velocity (right) at model locations where supercooled liquid water is predicted by the model. The 6 and 9-hour forecasts at both 3 and 13 km are shown for 17 June 2010.

In the second quantitative study we compared moderate or greater (MOG) icing PIREPs to the 3 and 13-km baseline wintertime model runs. Most icing occurs in winter, summer icing conditions are often short-lived, small-scale and usually occur in or near convection, which is discussed in the convective section of this document. Previous studies have suggested that in winter the Rapid Update Cycle (RUC) model significantly under-forecasts both the location and amount of supercooled liquid water relative to icing PIREPs (Wolff et al. 2010). In this study we mapped observations of MOG icing PIREPs, within 1 hour of the model valid time, to the sixteen closest 3-km WRF gridpoints and the closest 13-km gridpoint (Fig. 5.3). The top and base of the PIREP were determined from the report (often, pilots report icing through a layer during ascent and descent) and the associated gridpoints were extracted from the model runs. Only sub-freezing gridpoints mapped to a PIREP were used in the study. The lowest and highest temperatures, maximum relative humidity (RH), maximum vertical velocity (W), maximum supercooled cloud water plus supercooled rain water (SLW) and maximum ice phase condensate (sum of cloud ice, snow and graupel (ICE)) at all the gridpoints associated with the MOG PIREP were determined (Fig. 5.4). If any of the tests described below exceeded a test threshold then the condition was met.

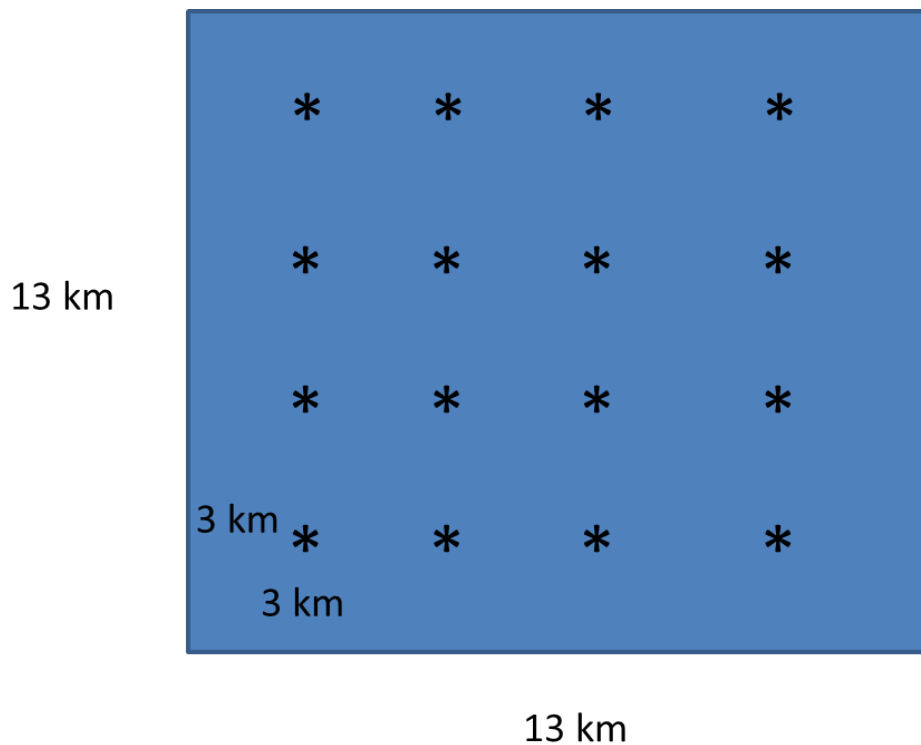


Figure 5.3: Areal coverage and associated gridpoints mapped to a co-located PIREP. Blue box represents the 13-km gridpoint and the black stars represent the 3-km gridpoints.

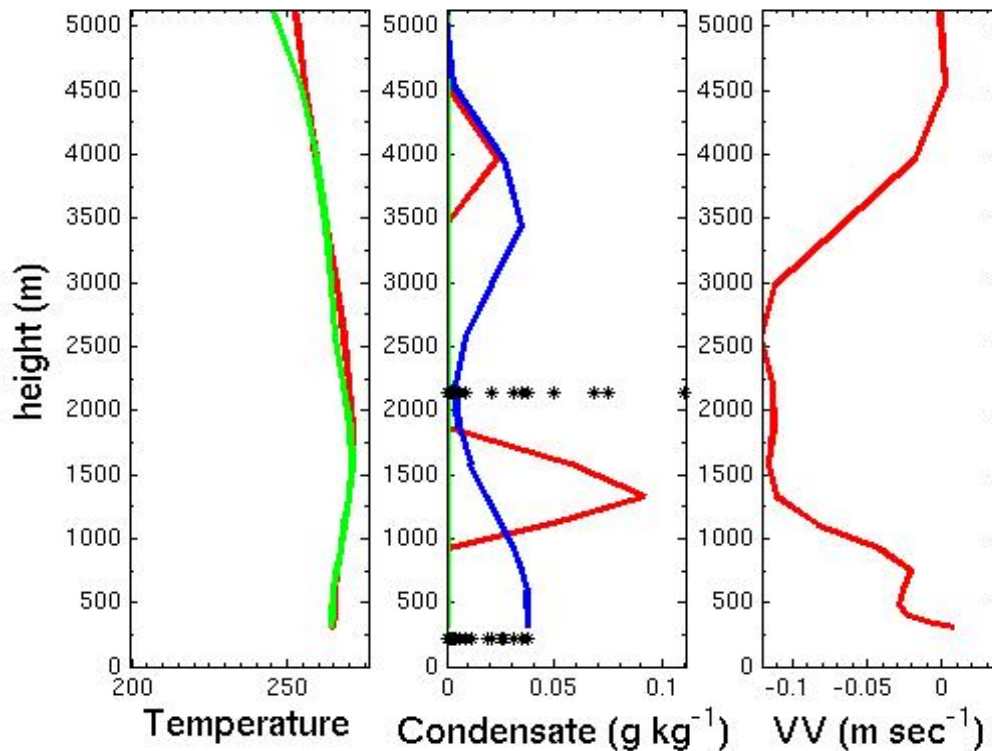


Figure 5.4: Vertical profiles of parameters from the 3-km WRF at a PIREP location. Temperature (red), dew point temperature (green) as a function of height shown on left, model condensate (red – SLW, blue – ICE), PIREP top and base as black stars) center, and vertical velocity on right.

5.1.1 PIREP vs. Model Tests

The 60 MOG icing PIREPs from the 1500, 1800, and 2100 UTC forecasts from 16 November 2010 and 1 February 2011 suggest that the 13-km WRF predicted SLW more often than the 3-km version (Table 5.1). Both the 3-km and 13-km versions were more likely to predict ICE than SLW at the PIREP locations. On average the 13-km WRF predicted a cloud (either ice or liquid) at 96% of the icing locations. The 3-km model predicted a cloud on average at 74% of the PIREP locations. Most of the model forecasts had model relative humidities exceeding 90% (icing occurs at $\text{RH} \geq 100\%$ in the atmosphere), although the 13-km was more moist than the 3-km. Almost all of the PIREPs occurred in rising air.

MOG PIREPs which occurred at temperatures colder than -12°C were also assessed. The results suggest that the model is producing ice in lieu of water at these temperatures, as the PODs for ice co-location are high and for SLW low.

Table 5.1: The results of the PIREP verification for six model forecasts over two winter days (16 November 2010 and 1 February 2011) for both the 3-km and 13-km WRF runs. The models were initialized at 1200 UTC on both days.

	2010 Nov 16 1500 UTC		2010 Nov 16 1800 UTC		2010 Nov 16 2100 UTC		2011 Feb 1 1500 UTC		2011 Feb 1 1800 UTC		2011 Feb 1 2100 UTC	
	13- km	3-km	13-km	3-km	13- km	3-km	13- km	3-km	13- km	3-km	13- km	3- km
POD SLW	33%	26%	0%	7%	13%	11%	100%	75%	100%	25%	20%	14%
POD ICE	61%	47%	100%	78%	87%	78%	0%	0%	0%	50%	60%	33%
POD RH >90	61%	26%	100%	86%	60%	44%	100%	75%	100%	50%	80%	50%
POD W – up	100%	100%	100%	93%	100%	100%	100%	100%	100%	100%	100%	67%
POD – C SLW	7.1%	0%	0%	0%	7%	0%	na	na	100%	25%	0%	0%
POD- C-ICE	92.5%	47%	100%	92%	93%	77%	na	na	0%	50%	100%	67%

POD SLW – percent of MOG PIREPs with any co-located model predicted supercooled liquid (cloud and/or rain water) $> 0.001 \text{ g kg}^{-1}$.

POD ICE - percent of MOG PIREPs with no co-located model predicted supercooled liquid but with frozen condensate (cloud ice, snow and/or graupel) $> 0.001 \text{ g kg}^{-1}$.

POD RH>90 – percent of MOG PIREPs with any co-located model predicted relative humidity $> 90\%$.

POD W-up – percent of MOG PIREPs with any co-located rising air.

POD C SLW - percent of MOG PIREPs that occurred at temperatures $< -12^{\circ}\text{C}$, and with any co-located model predicted supercooled liquid (cloud and/or rain water) $> 0.001 \text{ g kg}^{-1}$.

POD C ICE - percent of MOG PIREPs that occurred at temperatures $< -12^{\circ}\text{C}$, and with no co-located model predicted supercooled liquid but with frozen condensate (cloud ice and/or snow, and/or graupel) $> 0.001 \text{ g kg}^{-1}$.

5.2 Qualitative Analysis

The fine-scale quantitative analysis suggested the 3-km model was able to generate higher vertical velocities (positive and negative) and amounts of condensate than the 13-km model. The PIREP analysis suggested the 13-km model might have been slightly better at predicting icing, but before we make this somewhat critical conclusion we must look at more cases and/or deal with forecast scaling issues (i.e., is it appropriate to present a forecast with 1-hour granularity at 3-km resolution?). The larger-scale analysis suggested the models were fairly similar in the prediction and locations of condensate and vertical velocity. The northwest quarter of the domain at the 30th vertical level is used as an example (Fig. 5.5). The total condensate forecasts (all model condensate summed) are very similar in both models, although more wave-like cloud fields are often observed in the 3-km forecasts (Fig. 5.6). After reviewing the vertical structure of the model soundings at several icing locations (as in Fig. 5.4), a surprising number of the matched 3-km soundings were associated with descending air motions adjacent to model soundings with rising air motion. These up/down couplets are much more frequently observed in the 3-km WRF runs than in the 13-km model forecasts (Fig. 5.7). This may be expected due to the 13-km version's inability to resolve features of such small scale, but it will have an effect on icing forecasting.

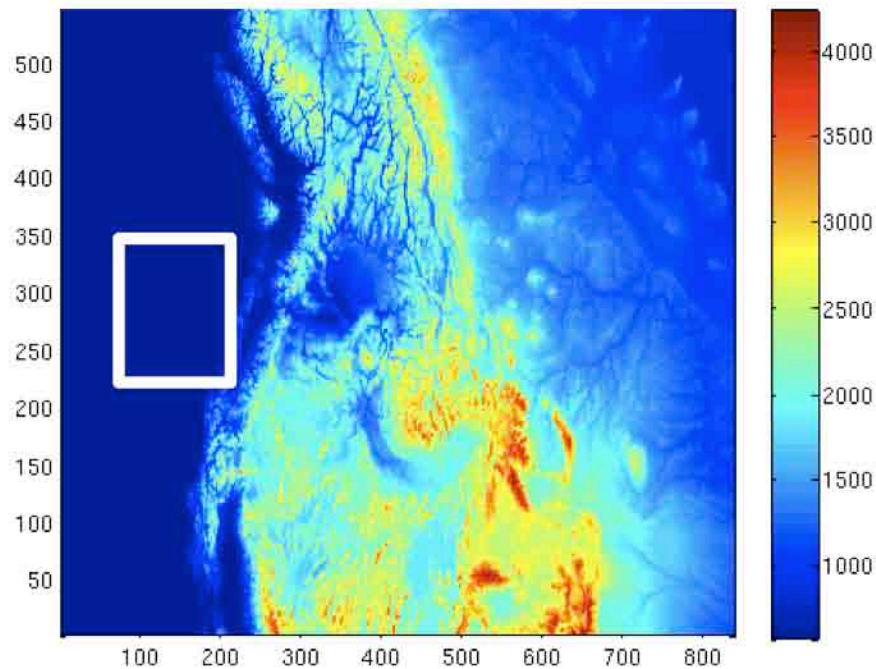


Figure 5.5: The Northwest quarter of the domain at the height (gpm) of the 30th vertical level (3-km WRF). The vertical and horizontal axes define the 3 km WRF gridpoints across the domain. The white box denotes sub-grid location.

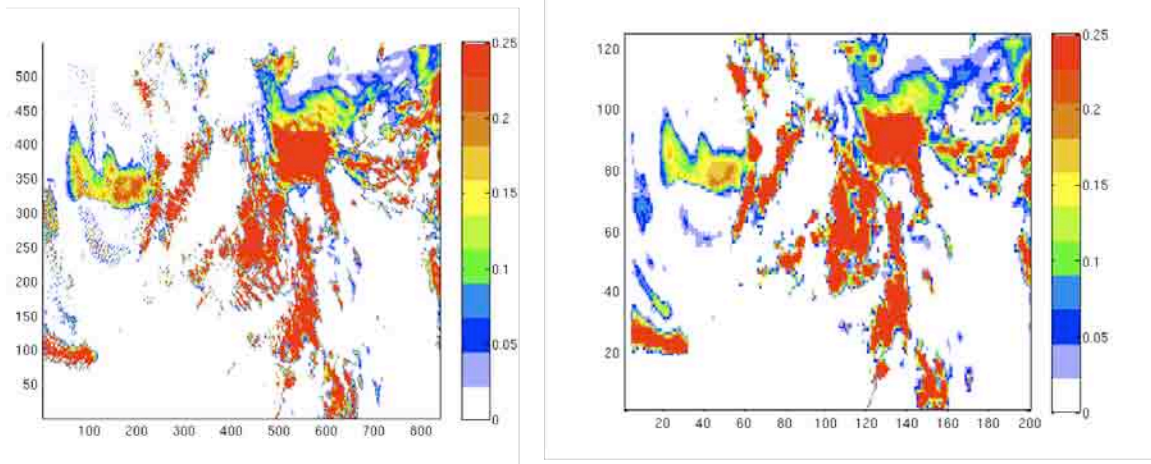


Figure 5.6: The model total condensate (g kg^{-1}) over the Northwest quarter of the model domain on 16 November 2010 at 1500 UTC at the $k = 30$ vertical level. The 3-km forecast is on the left and 13-km is on the right. The vertical and horizontal axes define the 3 km and 13 km WRF gridpoints across the domain.

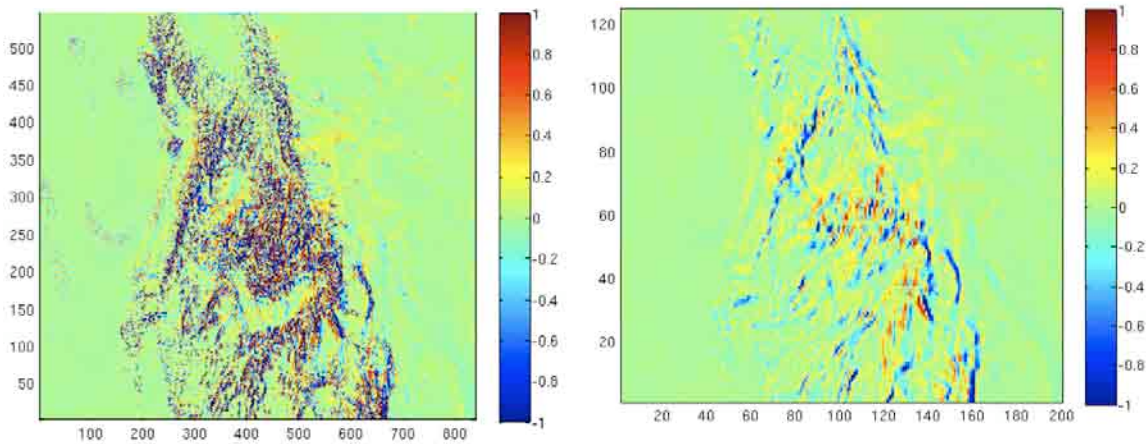


Figure 5.7: The vertical velocity (m sec^{-1}) over the northwest quarter of the model domain on 16 November 2010 at 1500 UTC at the $k = 30$ vertical level. The 3-km WRF forecast (left) and 13-km (right) are presented. The vertical and horizontal axes define the 3 km and 13 km WRF gridpoints across the domain.

To further analyze these waves, a sub-domain from the 3-km model was selected off the Oregon coast (see white box in Fig. 5.5). Comparison of the vertical velocity and total condensate fields shows what we expect: the rising air in the model predictions creates clouds at that scale. This suggests that the waves are not very short lived and the condensate has an opportunity to react to these waves. These waves, however, were not obvious in the 5 km GOES West IR image for 1430 UTC on 16 November 2010 (Fig. 5.9), thus questioning whether the model-predicted waves are indeed real? Either way, for prediction of in-flight icing conditions we need to take these high-resolution

features into account by averaging over larger scales appropriate to our forecast window (for example, our icing forecasts currently have 1-hour granularity).

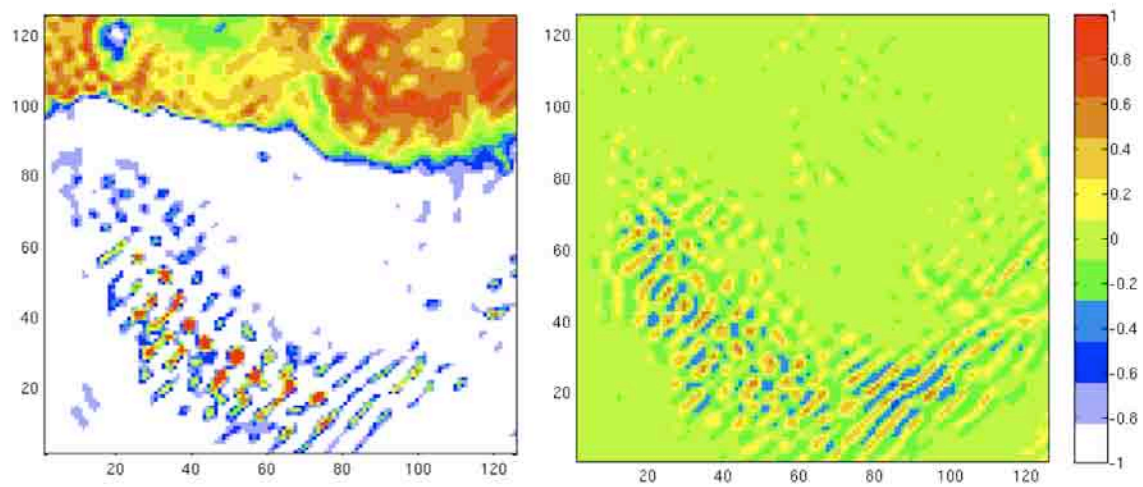


Figure 5.8: The 3-km model total condensate (g kg^{-1}) (left) and vertical velocity (m sec^{-1}) (right) waves over the Northwest sub-grid (white box in Fig. 5.5) on 16 November 2010 at 1500 UTC at the $k = 30$ vertical level. The vertical and horizontal axes define the 3 km WRF gridpoints across the sub-domain.

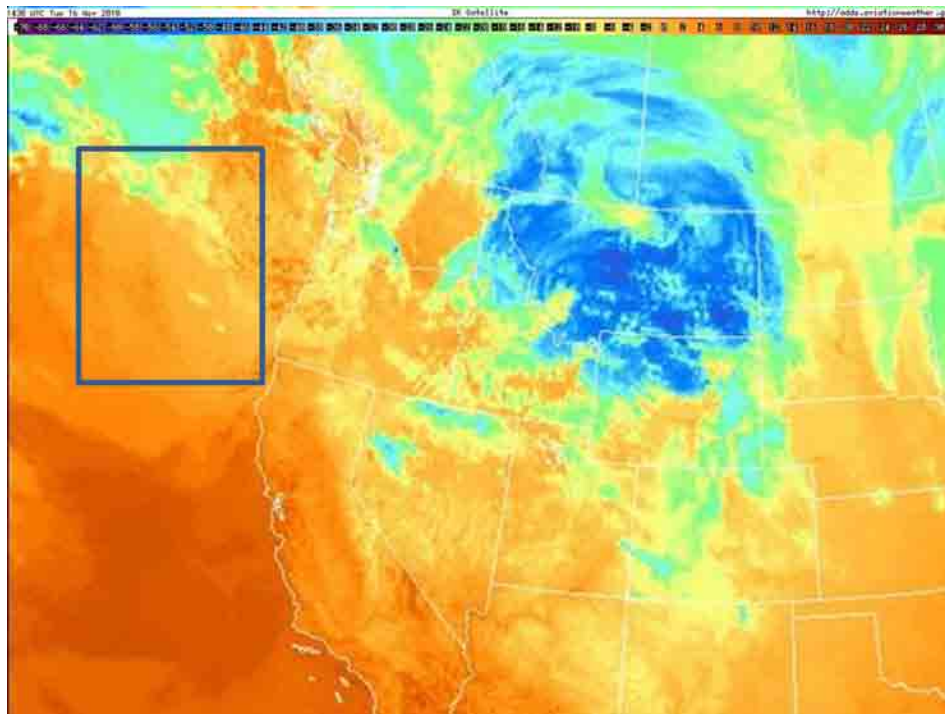


Figure 5.9: IR GOES image (5 km resolution) at 1430 UTC on 16 November 2010. The approximate northwest sub-grid is shown as a blue box.

5.3 Principal Findings

The 3-km WRF was able to generate higher (positive and negative) vertical velocities than the 13-km WRF. This occurred more often in the summer cases. The stronger updrafts in the 3-km WRF generated additional condensate compared to the 13-km WRF.

While both models did an excellent job predicting a model cloud (either ice or liquid) at the PIREP locations, they appeared to over predict ice and under-predict supercooled liquid water. This is based on extent of coverage, not quantity of condensate.

Neither model correctly identified icing conditions at temperatures colder than -12°C. They both correctly identified cloud (ice or liquid) but incorrectly predicted ice rather than supercooled liquid water the majority of the time.

Wave-like structures were observed across the 3-km domain. They appear to be associated with the predictions of condensate. These waves may be model artifacts or they may be real structures. If they are deemed significant in terms of forecasting aviation weather hazards, they should be investigated further.

References for Section 5

- Bernstein, B.C., McDonough, F., Politovich, M. K., Brown, B. G., Ratvasky, T.P., Miller, D.R., Wolff, C.A., and Cuning, G., 2005: Current Icing Potential: Algorithm Description and Comparison with Aircraft Observations. *J. Appl. Meteor.*, **44**, 969 – 986.
- Wolff, C.A., and McDonough, F., 2010: A comparison of WRF-RR and RUC forecasts of aircraft icing conditions. Paper 6.2, *14th Conference on Aviation, Range, and Aerospace Meteorology*, 23 – 26 January 2010, Atlanta, GA, available online.

6. Analysis focused on cold-season Surface Precipitation Coverage and Phase

6.1 Introduction

Snow removal of airport runways and de-icing procedures cause a majority of air traffic delays during the winter season. Accurate predictions of timing and coverage of cold-season precipitation type or phase reaching the ground is thus critical for air traffic operations. Here, we examine the ability of numerical weather prediction models to make accurate forecasts of surface precipitation phase and areal coverage by comparing the model output data with surface observations. Data from four baseline runs at 13 km (emulating the WRF-RR model) and 3 km (emulating the HRRR model) grid resolutions were analyzed. We selected two cases of relatively weak synoptic events (25 October 2010, 16 November 2010) and two additional cases of strong synoptic storms (4 February 2010, 1 February 2011). Qualitative and quantitative forecast skills are discussed in this report. In addition, comparisons of model performance between the 13-km and 3-km baseline runs were made. Model sensitivity to the microphysics parameterization was also investigated.

6.2 Dataset

6.2.1 Surface Observation

Surface observations used for verification come from the Automated Surface Observing System (ASOS). At over 800 ASOS sites throughout the U.S., precipitation identification sensors, known as Light Emitting Diode Weather Identifier (LEDWI), continuously record precipitation type every minute. Precipitation types identified by the sensor are rain, snow, and unknown precipitation (UP) depending on the particle terminal velocity. UP is an indication of occurrence of light precipitation, such as light snow and/or drizzle, but it is too light to be determined whether it is in solid or liquid phase. Raindrops detected by the sensor are categorized as freezing rain if the temperature is below 0°C and the collocated freezing rain sensor indicates an immediate freezing of the raindrops on the sensor's heated wire. According to the ASOS manual, snow pellets and snow grains are categorized as snow, but ice pellets may be falsely reported as rain due to their relatively high terminal fall velocity. For this study, the 1-minute ASOS data were quality controlled to catch suspect reports, such as blowing snow and UP due to haze/fog, by examining the temporal and spatial consistency in the observations and utilizing insights from other ASOS sensors (e.g., temperature and dew point, visibility).

6.2.2 Model Data

Surface precipitation types generated from the WRF model emulating HRRR with Thompson microphysics are rain, snow, and graupel. They are based on the instantaneous surface precipitation amount (e.g., the amount generated over the model time step of 18 seconds) for the three precipitating species valid at each forecast hour. The three precipitation types are combined to specify areas of snow, rain, and mixed-phase precipitation to compare with the surface observations. In this study, the model surface precipitation is defined as mixed-phase when both snow and rain are present in the grid cell **and** the snow water mixing ratio is less than

60% of the total precipitation mixing ratio. The definition of mixed-phase precipitation used in this study is the same as in the operational HRRR model.

The simulations with a 13-km grid resolution were performed with the Grell-Devenyi ensemble cumulus parameterization scheme (Grell and Devenyi, 2002). It uses a multi-closure, multi-parameter, ensemble method to express the interaction between the convection and ambient forcing. The WRF model stores surface precipitation data from the cumulus scheme separately from the explicit (microphysics) scheme. Thus precipitation comes from both parameterized and explicit portions. Generally, contribution from the cumulus parameterization is relatively small in cool-season compared with warm-season precipitation. Unlike the explicit microphysics module, precipitation from the cumulus parameterization module does not differentiate between various precipitation types. For this study, the parameterized precipitation was separated into snow and rain categories simply based on surface temperature. For the cases examined, precipitation from the cumulus parameterization predominantly impacted rain regions.

6.2.3 Verification Approach

The “truth” data from the 1-minute ASOS information was generated by taking the observed type within a temporal window centered at each hourly forecast time. When a rain-to-snow transition or freezing rain occurred during the temporal window, the “truth” was classified as mixed-phase. If snow was the only reported type or snow and UP occurred during the time period but no rain, the most frequently reported of the two types was chosen as the “truth”. Assigning rain as the “truth” was based on the same criteria. Note that UP was used to verify the presence of precipitation but not of any particular precipitation phase.

Categorical (dichotomous) verification statistics were computed by matching the model precipitation phase with the “truth” inside a neighborhood box centered at each ASOS site. Verification statistics were determined for 1) areal coverage of all precipitation types combined, 2) areal coverage of snow, 3) rain, and 4) mixed-phase. Verification measurements used to assess model performance in this study are frequency bias, probability of detection (POD), false alarm ratios (FAR), critical success index (CSI), probability of false detection (POFD), and probability of null detection (PODn).

Table 6.1: a) temporal window sizes centered at forecast valid time and b) neighborhood box size examined in this study.

<i>(a) Temporal window size centered at forecast valid time</i>				
<i>+/- 3min</i>	<i>+/- 6 min</i>	<i>+/- 10 min</i>	<i>+/- 30 min</i>	<i>+/- 60 min</i>
<i>(b) Neighborhood box size (km)</i>				
3-km model	12	24	48	
13-km model		26	52	

The temporal window and neighborhood box sizes tested for sensitivity are listed in Table 6.1. Quantitative assessments with various neighborhood box sizes and temporal window sizes are presented in the result section.

6.3 Analysis Results

6.3.1 Examination of Precipitation Data Field Masking

The numerical forecast models can generate precipitation amount much less than most of the surface observational systems because of the high computer precision. The minimum precipitation rate the ASOS precipitation identification sensor can detect is ~ 0.01 inches per hour or $\sim 7 \times 10^{-5} \text{ mm s}^{-1}$ ($\sim 1.3 \times 10^{-3} \text{ mm/18-sec time-step}$). Therefore, a threshold was applied to the model precipitation field using the *time-step* precipitation amount. The question was “what is an appropriate threshold amount for cold-season precipitation?” In this section, we explore various data masks and examine associated changes in skill scores.

Table 6.2 lists various time-step precipitation amounts used as minimum threshold applied to the model precipitation data. The values indicated are in units of mm per time-step (18 seconds). Mask levels 1 – 3 use a constant threshold value on the rain, snow, and graupel fields. Thresholds applied to the three precipitation types for mask levels 4 – 5 are not the same. The thought here was to vary the threshold amount with respect to the rainwater density. Specifically, the thresholds on snow and graupel fields are 10 and 40% of that of the rain field, respectively, based on the snow-rain and graupel-rain density ratios typically used in a microphysics scheme.

Table 6.2: List of threshold on the model precipitation field. The values are in mm per time-step (18 seconds). Under Mask Type, (c) indicates that the same threshold values are used for the three species, and (v) is for using different threshold.

<i>Mask Type</i>	<i>Rain</i>	<i>Snow</i>	<i>Graupel</i>
1 (c)	1e-6	1e-6	1e-6
2 (c)	1e-5	1e-5	1e-5
3 (c)	1e-4	1e-4	1e-4
4 (v)	1e-4	1e-5	4e-5
5 (v)	1e-3	1e-4	4e-4

Figure 6.1 shows spatial distributions of model predicted precipitation phase field valid at 23 UTC on 1 February 2011. Overlaid in Figs. 6.1a – d are validations for an occurrence of precipitation made at the individual ASOS sites. Observed precipitation phases at the surface sites are in Fig. 6.1e.

Verification valid at 1 February 2011 2300 UTC

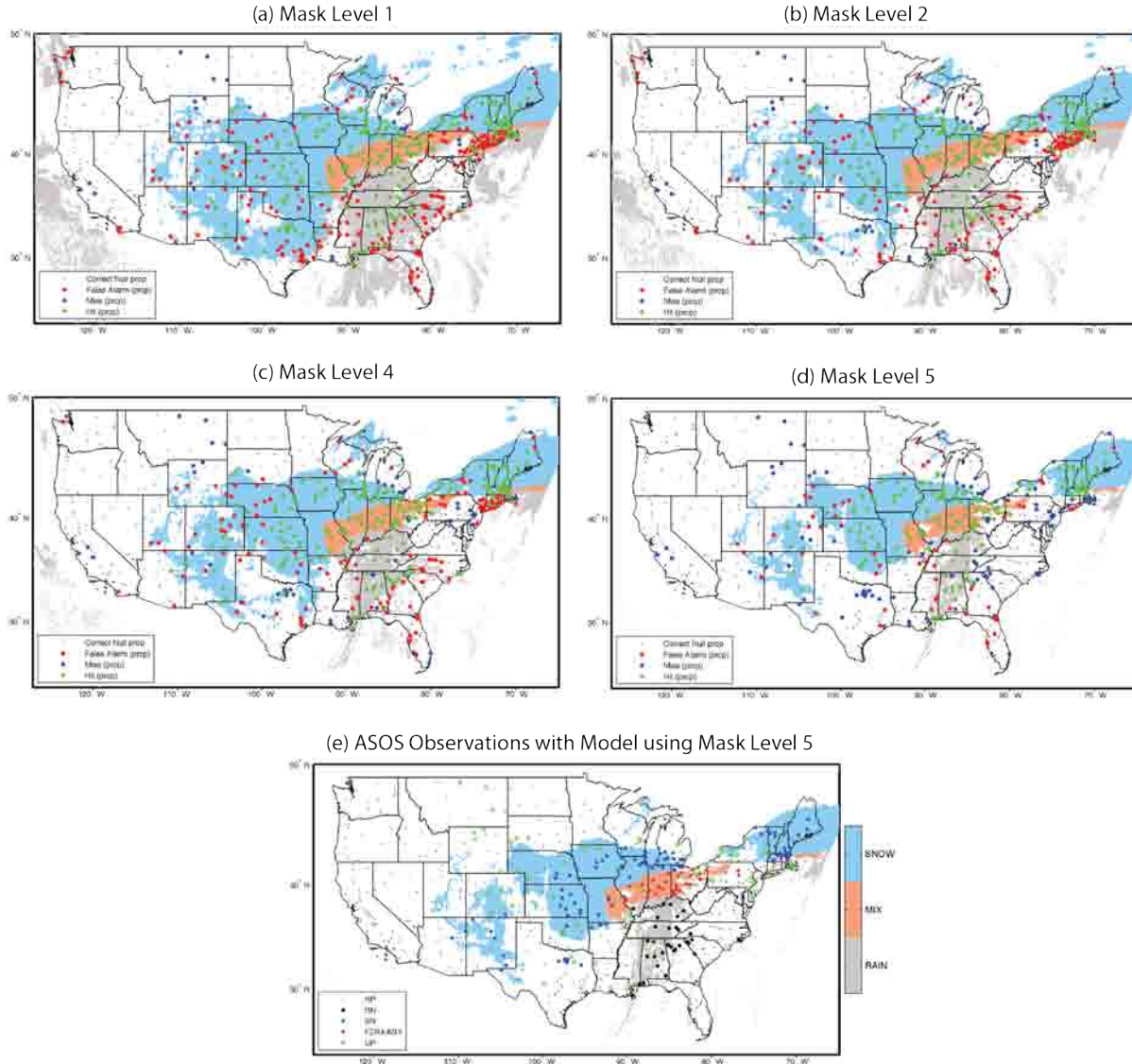


Figure 6.1: (a) – (d) Verification for the presence of precipitation valid at 23 UTC, 1 February 2011 (11 hour forecast) using masking level 1, 2, 4, and 5. (e) ASOS observations valid at model valid time with the mask level 5 data in the background.

With all masking levels the model produced a wide swath of mixed-phase precipitation as reported by ASOS. However, differences in the number of false alarms along the edges of the storm are quite significant among the various mask levels. With almost no data masking (Fig. 6.1a), there are a significant number of false alarms in the northeastern, southern, and western portions of the storm, clearly indicating that application of some threshold is indeed necessary. With mask level 2, the number of false alarms is reduced, but a fair number of them remain in the southeast U.S. (Fig. 6.1b). Mask level 4 has an order of magnitude higher threshold level for rain while the threshold level for snow is the same as in mask level 2. With the set of threshold

values for mask level 4, the comparison between the model and observations in the southeast U.S. improves (Fig. 6.1c). With an order of magnitude higher value of thresholds for all species compared with mask level 4 (i.e., mask level 5), the agreement becomes even better (Fig. 6.1d).

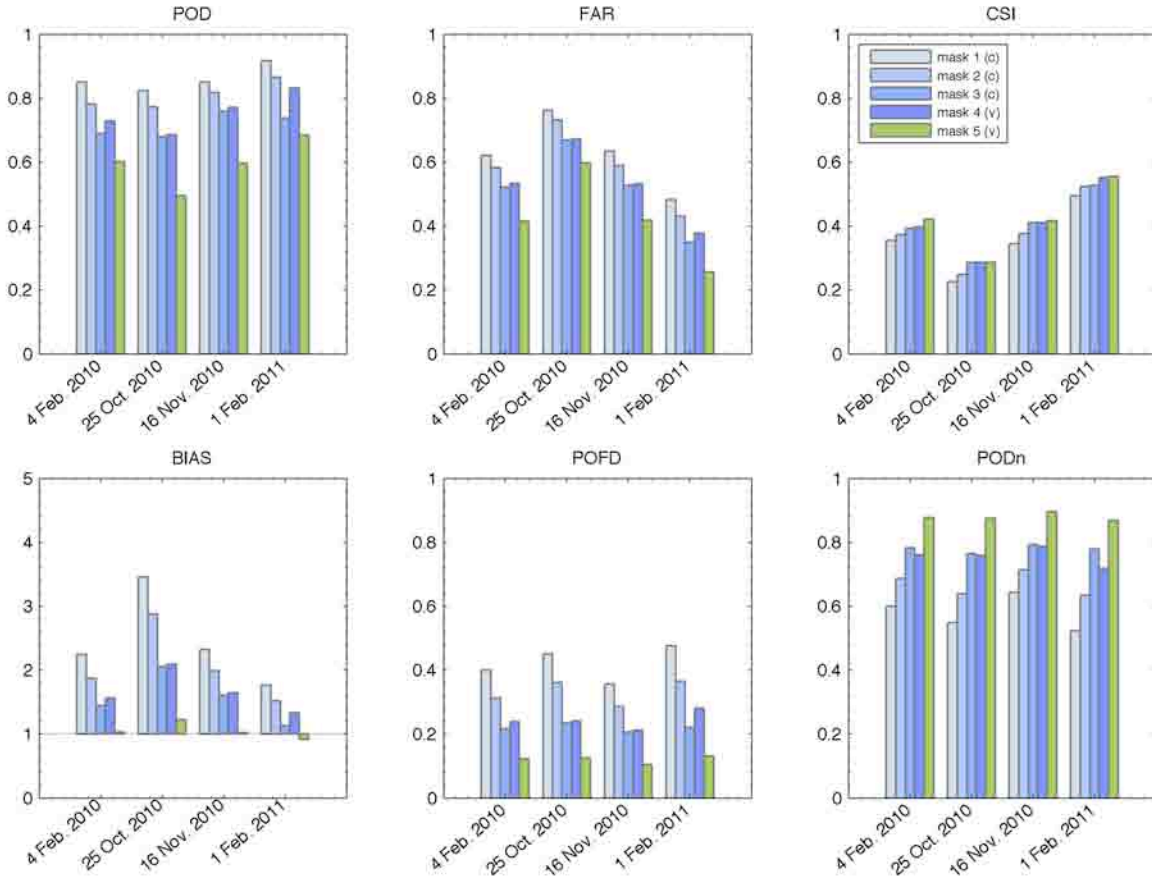


Figure 6.2: Verification measurements using various data masks for the four 3-km baseline runs examined in this study. The skill scores in this figure were computed using a neighborhood box width of 24 km and temporal window length of 12 minutes.

Figure 6.2 shows verification results for precipitation coverage from the four 3-km baseline runs. The verification measurements were computed from all 12 forecast lead times combined. Overall, all verification measurements show improvements as the mask level goes from 1 to 5. In particular, among the five mask levels tested, the most optimal results in bias and FAR are achieved with mask level 5. Consistently, CSI is the best with mask level 5. Both PODF and PODn also indicate that mask level 5 agrees better with the observations of “no precipitation.”

A fine-tuning of the threshold amount would require evaluating a larger number of cases. Nevertheless, this exercise has revealed that the precipitation mask based on a set of threshold values scaled by the snow-rain density and graupel-rain density ratios will give better correspondence with surface observations. All verification results presented hereafter are based on mask level 5.

6.3.2 Qualitative Analysis

Figures 6.3 – 6.6 compare model precipitation and observations valid at select forecast times from the four 3- and 13-km baseline runs. The forecast times selected were chosen to show representative features of the individual storms.

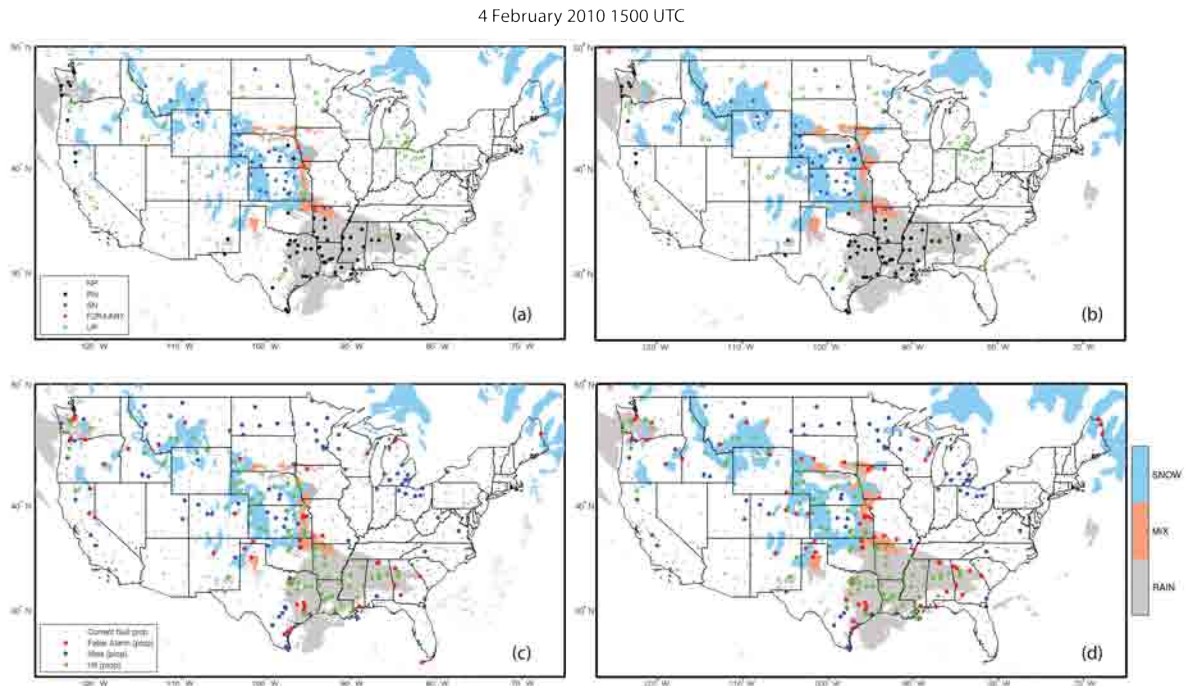


Figure 6.3: Precipitation field valid at 15 UTC 4 February 2010 from (a), (c) the 3-km baseline run and (b), (d) 13-km baseline run. Overlaid in panels (a) and (b) are ASOS observations at model valid time with a temporal window length of 12 minutes. On panels (c) and (d), validation (i.e., hit, false alarm, miss, correct null) compared with the ASOS observations are indicated by the circles. Model runs started from 12 UTC.

Case (1): 4 February 2010

A synoptic scale frontal system produced snow between the northern and central Great Plains and rain mainly in the Southern U.S. on 4 February 2010. Freezing precipitation was reported prior to snow arriving in the central Great Plains. The areal coverage of mixed-phase precipitation continuously formed a complex shape as the system evolved.

In general, the observed and model predicted precipitation phase agreed well in both the 3- and 13-km baseline runs (Figs. 6.3a – b). Regions where the model-predicted mixed-phase

precipitation swept across were generally correct in comparison with the time history of observations (not shown). Rain coverage is well represented by the model. At both model resolutions, light snow (indicated by green open circles) between Minnesota and Ohio were not predicted.

Figures 6.3c – d show the validation of precipitation occurrence at model valid time from the 3- and 13-km runs, respectively. Compared with the 3-km runs, there are more false alarms in the 13-km run.

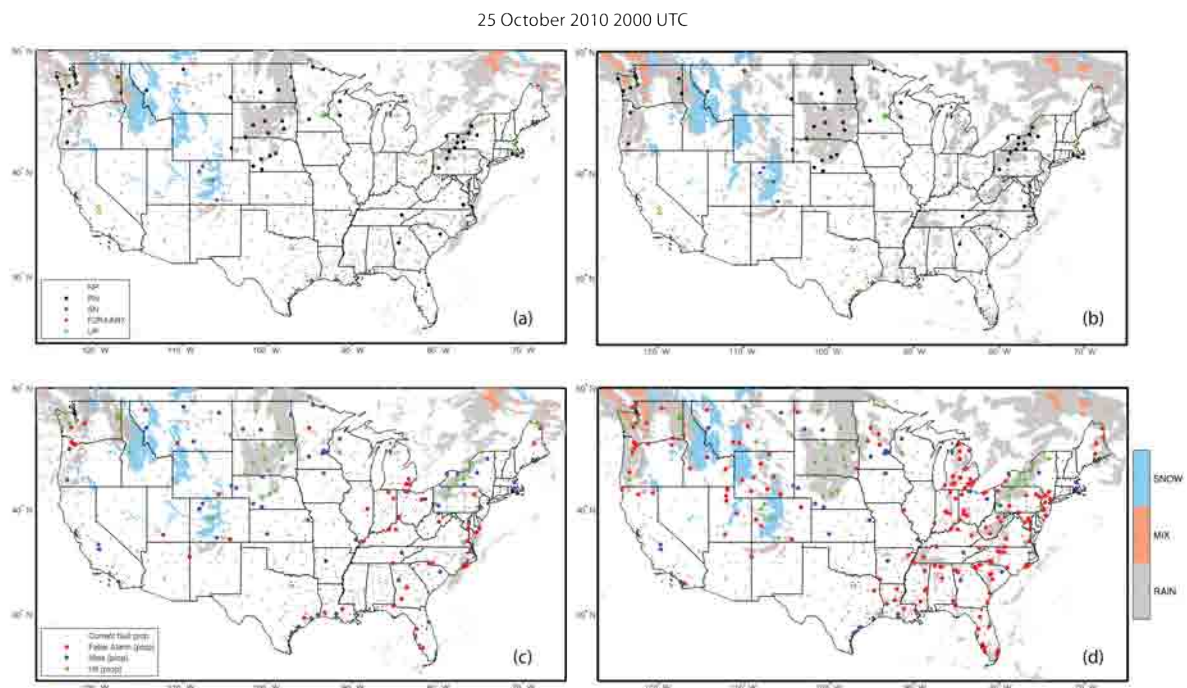


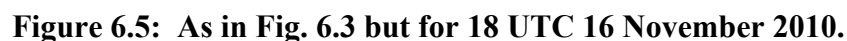
Figure 6.4: As in Fig. 6.3 but for 20 UTC 25 October 2010.

Case (2): 25 October 2010

The precipitation system on 25 October 2010 was characterized by light snow and rain in the Pacific Northwest and northern Great Plains from a weak upper level trough. (This statement is from a winter weather perspective. Convection occurred in the Southeast mainly before 20 UTC).

The number of ASOS sites in the Northwestern U.S., in particular, over the Cascade mountain range and the complex terrain of the Intermountain West is rather limited. In addition, ASOS sites are typically located in the valleys. Thus, as seen in Fig. 6.4a – b, surface observations to validate snow in the model are scarce. Nevertheless, snow seems to be predicted in a general

There are more false alarms in the 13-km run compared to the 3-km run (Figs. 6.4c-d), particularly in the Southeast. This result comes from heavier rainfall and numerous convective cells simulated from the cumulus parameterization in the 13-km model.



The 16 November 2010 case featured two separate systems. A convective rain system was in the Eastern U.S. and a winter precipitation system in the northern Great Plains.

Overall, precipitation coverage is well represented by the model at the two resolutions (Figs. 6.5a – b). The rain region in the Eastern U.S. matches well with the observation. As in the 25 October 2010 system, verifying snow coverage and rain-snow transition zone are complicated by the small number of ASOS sites in the region. Nevertheless, the time history of model – observation comparison (not shown) indicated that transient snow and a rain-to-snow transition occurred over the region.

Comparison between the 3- and 13-km runs indicates more false alarms in the rain region of the 13-km run. As for the 25 October 2010 case, this was associated with the area of convective rain produced from the cumulus parameterization.

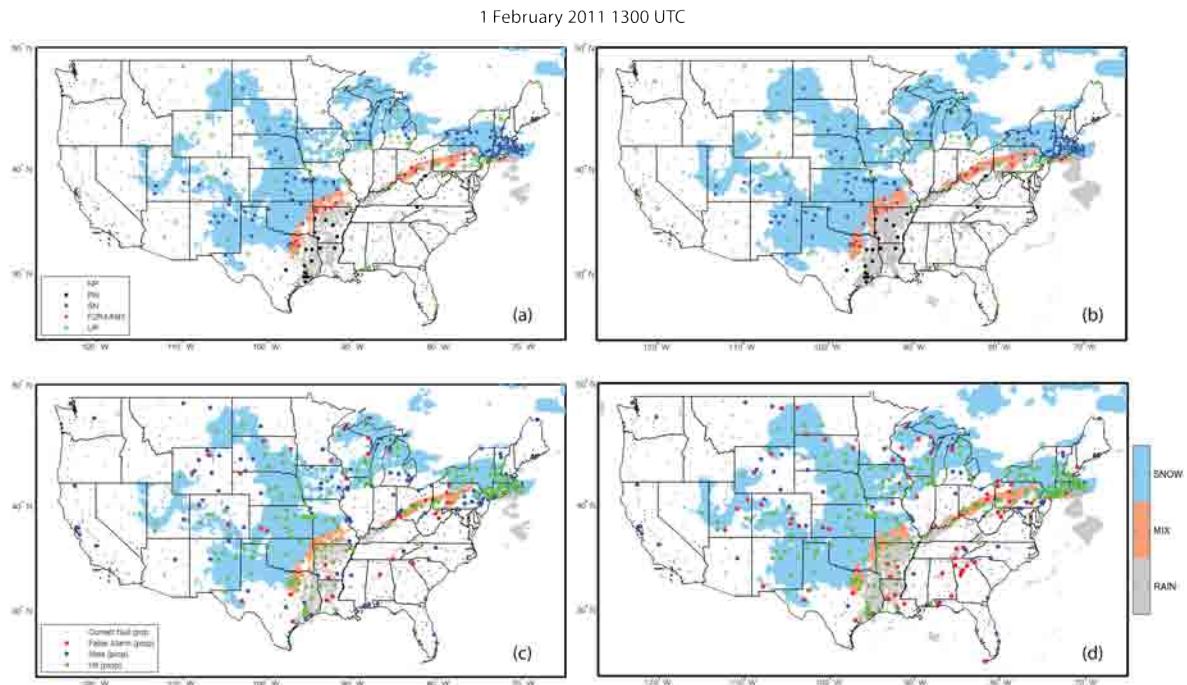


Figure 6.6: As in Fig. 6.3 but 13 UTC 1 February 2011.

Case (4): 1 February 2011

The storm on 1 February 2011 produced freezing rain over a widespread area from northeastern Texas all the way to the upper northeast coast, heavy snow in the northern part of the system and rain in the southeastern U.S. Convective activity within the storm system was suggested based on the surface [human] reports of thunder snow in Texas and graupel in Missouri and Iowa.

At both grid resolutions, the model performed very well in predicting the spatial distribution of each precipitation phase (Figs. 6.6a – b). The rain-snow transition zone is well represented by the model at both grid resolutions. Not surprisingly, the rain-snow transition zone is more broadly defined in the 13-km run compared to the 3-km run. Overall, the verification results for the presence of precipitation (Figs. 6.6c – d) show similar results in both runs. An exception is in Iowa, where the 3-km run shows more detailed features associated with precipitation that is thinning out. Over this region, the number of misses is larger in the 3-km run. Another area of difference between the 3- and 13-km models is in the Southeast. Over this particular region, the

cumulus parameterization in the 13-km model generated slightly more convective cells that resulted in more false alarms.

6.4 Quantitative Analysis

6.4.1 Forecast Skill Scores from the 3-km Baseline Runs

Figure 6.7 shows the frequency bias plotted as a function of forecast lead time for (a) precipitation coverage, (b) snow, (c) rain, and (d) mixed-phase precipitation regions from the four 3-km baseline runs. Bias is based on a neighborhood box width of 24 km and temporal window length of 12 minutes.

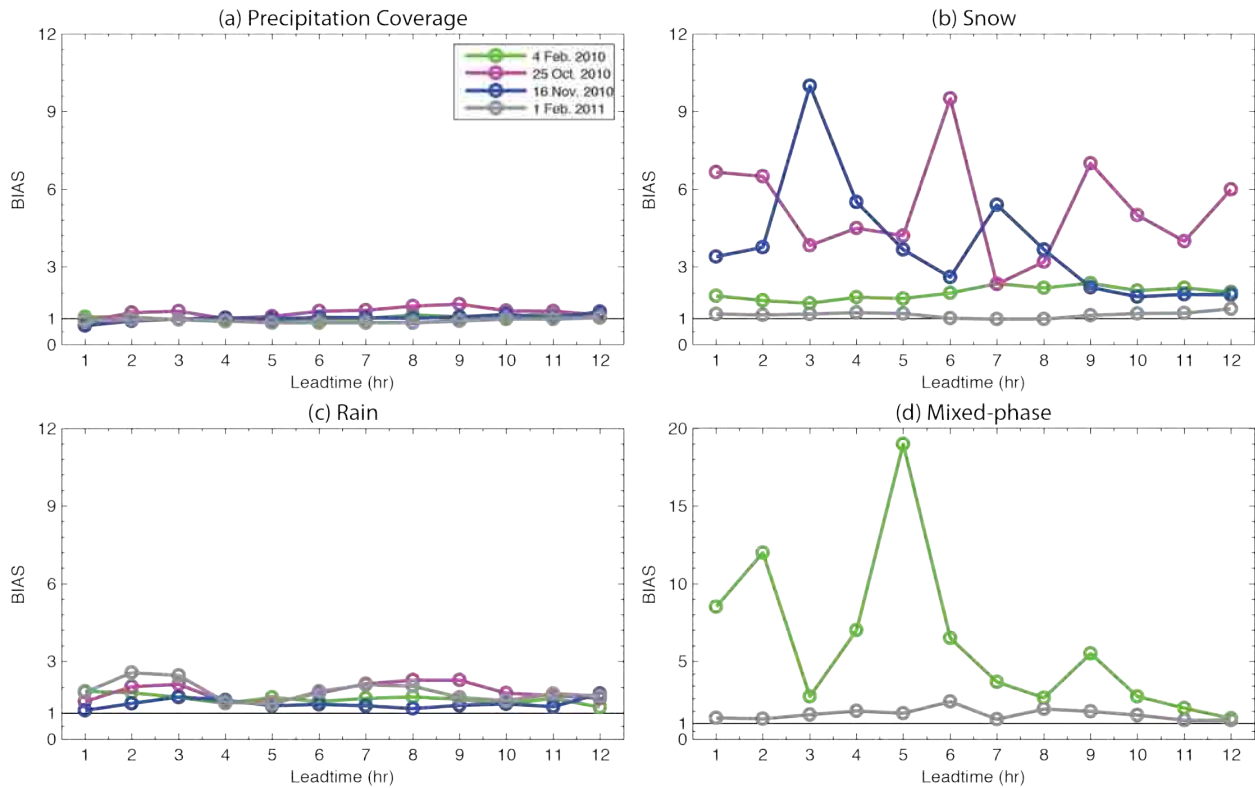


Figure 6.7: Forecast bias plotted as a function of lead time for (a) precipitation coverage, (b) snow, (c) rain, and (d) mixed-phase precipitation from the four 3-km baseline runs examined. Note that y-axis range is different from others. The values are from a neighborhood box width of 24 km and temporal window length of 12 minutes.

There is no significant trend with lead time in bias for precipitation coverage in all four cases (Fig. 6.7a). This is because these weather systems were of a synoptic scale and evolved slowly compared to warm-season convective systems. The bias averaged over the 12 lead times is between ~0.9 and 1.3 with a small hour-to-hour variation (Table 6.3), demonstrating good performance for areal coverage of precipitation in these cases. Similarly, the bias for rain coverage is reasonable with small hour-to-hour variation in all cases (Fig. 6.7c, Table 6.3).

Table 6.3: Average skill scores from 12 forecast lead times for (a) precipitation coverage, (b) snow, (c) rain, and (d) mixed-phase precipitation. Values in parentheses are one standard deviation from the mean values. The values are from a neighborhood box width of 24 km and temporal window length of 12 minutes.

<i>(a) Precipitation coverage</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	1.04 (0.09)	1.24 (0.18)	1.03 (0.12)	0.92 (0.06)
POD	0.61 (0.04)	0.50 (0.06)	0.60 (0.07)	0.68 (0.05)
FAR	0.42 (0.03)	0.59 (0.05)	0.42 (0.04)	0.26 (0.03)
CSI	0.42 (0.02)	0.29 (0.03)	0.42 (0.04)	0.55 (0.04)
<i>(b) Snow</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	2.00 (0.24)	5.23 (1.90)	3.82 (2.22)	1.15 (0.11)
POD	0.81 (0.10)	0.77 (0.29)	0.78 (0.14)	0.84 (0.05)
FAR	0.59 (0.04)	0.84 (0.07)	0.75 (0.10)	0.57 (0.05)
CSI	0.37 (0.05)	0.15 (0.06)	0.26 (0.08)	0.64 (0.05)
<i>(c) Rain</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	1.55 (0.17)	1.83 (0.32)	1.37 (0.18)	1.84 (0.37)
POD	0.76 (0.10)	0.61 (0.08)	0.75 (0.06)	0.78 (0.09)
FAR	0.51 (0.06)	0.66 (0.06)	0.45 (0.05)	0.57 (0.07)
CSI	0.43 (0.07)	0.28 (0.04)	0.47 (0.05)	0.38 (0.07)
<i>(d) Mixed-phase precipitation</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	6.14 (4.90)			1.59 (0.34)
POD	0.75 (0.28)			0.78 (0.11)
FAR	0.82 (0.10)			0.49 (0.11)
CSI	0.16 (0.10)			0.45 (0.11)

For areas of snow alone, the model bias is on the order of 1 – 2 for the 4 February 2010 and 1 February 2011 cases—two strong synoptic cases (Fig. 6.7b, Table 6.3). The average bias is significantly larger for the two weakly forced cases over the northern Great Plains (25 October

2010 and 16 November 2010) (Table 6.3). Additionally, there is more hour-to-hour variation compared to the strong synoptic cases. As mentioned in the previous section, this is because of the complexities of forecasting and validating the model output over the western U.S. Some of the complexity factors come from: 1) a limited number of ASOS sites, 2) the locations of ASOS sites being in valleys, b) elevation difference between model grids and observation sites, and c) the presence of local terrain induced effects producing transient precipitation shadows and rain-snow transition regions.

The bias for the mixed-phase precipitation region was computed only for the two strong synoptic cases. The number of surface observations was not enough to compute meaningful verification statistics for the two weakly forced cases. Bias for the 1 February 2011 case is nearly perfect. The bias is significantly larger for the 4 February 2010 case. This is not surprising because the mixed-phase region was less broadly defined and had a complex shape compared with the 1 February 2011 case (compare Figs. 6.3 and 6.6). Examination of the 4 February 2010 data indicated that false alarms were due to temporal offsets in the rain-snow transition that occurred at the leading edge of the main precipitation region in Iowa and eastern Dakotas. Accurate forecasting of the mixed-phase precipitation is challenging because various surface conditions, such as wet-bulb temperature and winds, impact precipitation phase during the cold-season.

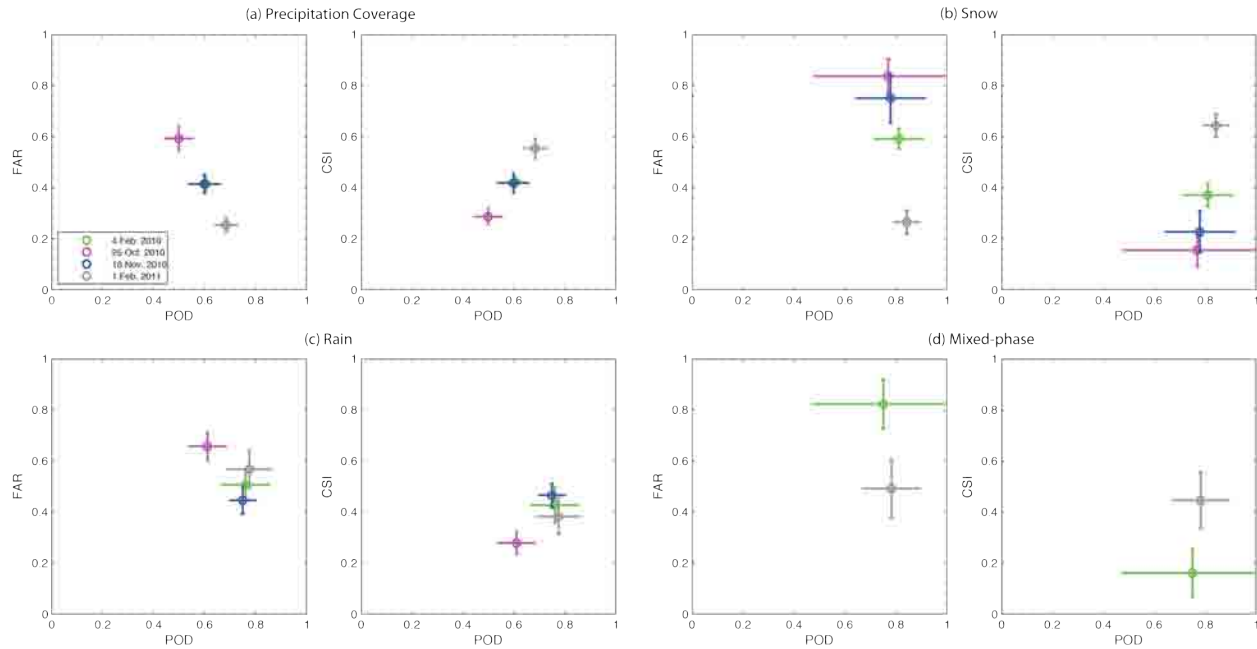


Figure 6.8: Scatterplots of FAR vs POD and CSI vs POD for (a) precipitation coverage, (b) snow, (c) rain, and (d) mixed-phase precipitation from the 3-km baseline runs. Colored circles are the average values over the 12-hr forecast and vertical and horizontal lines indicate one standard deviation from the mean. For (d), only data from 4 February 2010 and 1 February 2011 cases are shown because the other two cases did not have enough observations of mixed-phase precipitation. The values are from a neighborhood box width of 24 km and temporal window length of 12 minutes.

Scatter plots of 12-hour average FAR, POD, and CSI are shown in Fig. 6.8. The horizontal and vertical lines are one standard deviation from the average value (open circles). These values are tabulated in Table 6.3. In the FAR vs POD graphs, points of a perfect forecast would fall in the lower right corner; whereas in the CSI vs POD graphs, they would be in the upper right corner.

The scatter plots reflect the results shown in Fig. 6.7. In general, the better defined (or more widespread and strong) the precipitation systems are the higher the model's performance scores. For example, the relatively minor skill for the 25 October 2010 event has to be attributed to false alarms and misses associated with the transient small rain bands and convective cells in the southeastern U.S. and spotty snow areas over the northern Great Plains (Fig. 6.4). The ASOS sites over these regions indeed observed such small-scale precipitation features. However, any temporal and spatial offsets of such transient features compared to the observations affected the skill scores.

6.4.2 Comparison of Model Performance between the 13 and 3 km Runs

Table 6.4 shows the mean difference of the 12 individual hourly bias and CSI scores between the 13 and 3 km runs. The values in parentheses are the standard deviation of the mean differences. Bold fonts are used to indicate the mean difference being statistically significant, i.e., the mean bias and CSI from the 13-km runs are more than one standard deviation away from those of the 3-km runs (Table 6.3). The neighborhood box width used for the comparison is 24 and 26 km for the 3- and 13-km runs, respectively. The temporal window was 12 minutes.

From the perspective of surface precipitation type, the difference between the 13- and 3-km grid resolutions mainly appeared in the rain areas where the cumulus parameterization of the 13-km runs was most active. The active cumulus scheme produced a relatively large number of false alarms particularly in the southern and southeastern U.S. (warm and moist regions) compared to the 3-km runs (e.g., Figs. 6.4 – 6.6). As a result, the mean biases for the rain coverage were significantly larger and deviated away from unity (bias = 1) in the 13-km runs. These values were more than one standard deviation away from the mean bias in the 3-km runs (Table 6.3c). The large number of false alarm also affected the CSI scores in the rain areas (Table 6.4c) and the mean bias for the overall precipitation coverage (Table 6.4a).

Over the snow areas (Table 6.4b), the mean differences were small and most of them were within one standard deviation of the corresponding mean bias and CSI listed in Table 6.3b. This is not surprising because cumulus parameterization have little impact over snow regions where, in general, large-scale forces dominate.

The values in bold font in Table 6.4d are for the mixed-phase precipitation on 1 February 2011. In this case, the 13-km run produced a larger bias compared to the 3-km run due to more false alarms along the perimeter of the broad rain-snow transition zone. This is simply attributed to the larger grid size making the mixed-phase region broader. Because of the limited number of cases examined, the grid resolution impact on skill in forecasting the mixed-phase precipitation coverage is difficult to gauge any further.

Table 6.4: The mean difference in hourly bias and CSI between the 13- and 3-km base runs. The values in prentices are the standard deviation from the mean difference. A neighborhood box width of 24 and 26 km were used for the 3- and 13-km runs, respectively, to compute the verification statistics. In both datasets, a temporal window length of 12 min was used. Values in bold fonts are those where the difference is larger than one standard deviation of the mean values listed in Table 6.3.

<i>(a) Precipitation coverage</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	0.11 (0.03)	0.48 (0.27)	0.35 (0.10)	0.19 (0.03)
CSI	0.02 (0.02)	-0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
<i>(b) Snow</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	0.06 (0.08)	1.06 (2.46)	0.36 (0.59)	0.19 (0.03)
CSI	0.01 (0.02)	0.02 (0.07)	0.02 (0.06)	-0.02 (0.02)
<i>(c) Rain</i>				
	4 Feb 2010	25 Oct 2010	16 Nov 2010	1 Feb 2011
Bias	0.14 (0.13)	0.88 (0.53)	0.58 (0.23)	0.45 (0.30)
CSI	-0.01 (0.03)	-0.04 (0.03)	-0.05 (0.04)	-0.07 (0.04)
<i>(d) Mixed-phase precipitation</i>				
	1 Feb 2010	25 Oct 2010	16 Nov 2010	4 Feb 2011
Bias	-0.82 (3.57)			0.47 (0.30)
CSI	0 (0.08)			-0.05 (0.05)

6.4.3 Sensitivity to neighborhood box size and temporal window

Skill scores for assessing model performance vary with neighborhood box size and temporal window length. In order to examine change in skill scores with various box sizes and temporal window lengths, verification measurements were computed for all combinations of box size and temporal window length listed in Table 6.1.

Up to this point, the results computed from a neighborhood box size of 24 km and temporal window length of 12 min were shown. Using them as “control” verification statistics, deviations from them were computed for all other combinations of box size and temporal window length. The deviations from the “control” bias and CSI for verification of precipitation coverage are shown in Fig. 6.9. In general, the bias increases with box size and temporal window length in all cases. However, bias changes more sensitively with box size than with temporal window length. In fact, the amount of change in bias produced from varying the temporal window length is within one standard deviation of the 12-hour mean bias from the “control” (Table 6.3), indicating

that the differences are not statistically significant. On the other hand, the difference in bias among various box sizes is larger than one standard deviation from the 12-hour mean bias from the “control”.

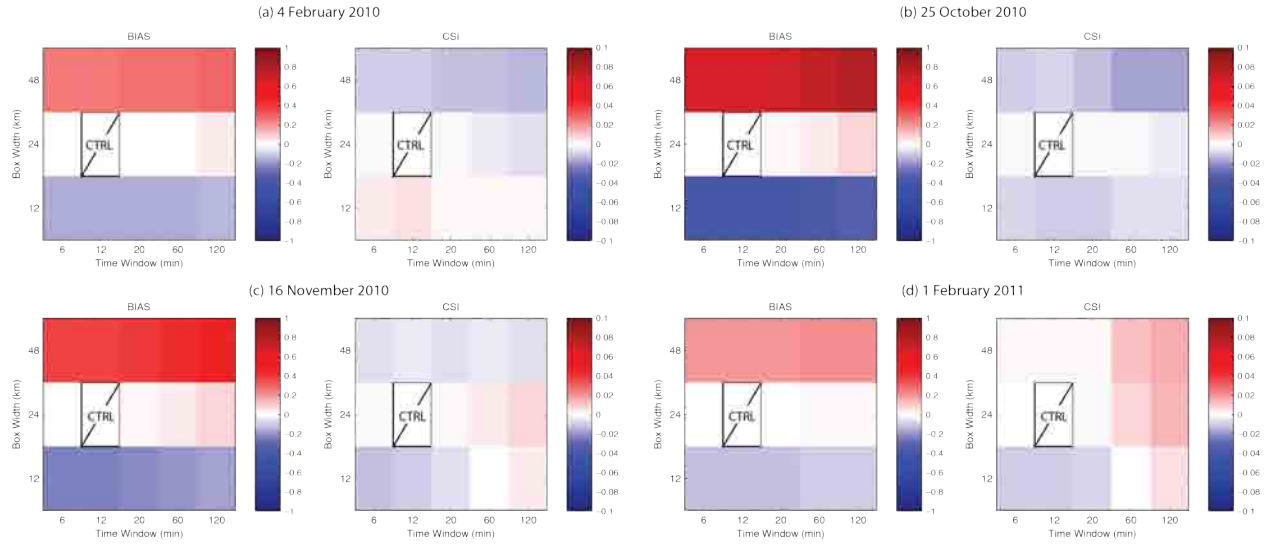


Figure 6.9: Spectral plots showing deviation of average bias and CSI from the control neighborhood box size and temporal window length from (a) 4 February 2010, (b) 25 October 2010, (c) 16 November 2010, and (d) 4 February 2011. The control box size and temporal window length are 24 km and 12 minutes, respectively (indicated as CTRL in the figure).

The deviations in CSI from the “control” show case-to-case variation. For the 4 February 2010 case, CSI decreases with box size and temporal window length. However, it decreases as box size and temporal window size shrink in the 1 February 2011 case. For the 25 October 2010 case, CSI is maximized with the 24-km and 12-min combination. Regardless of such different behaviors, comparisons with standard variations listed in Table 6.3 reveal that the changes are all within one standard deviation from the average CSI from the “control”, indicating that the CSI differences are not so significant. This is due to a compensating effect of gains/losses in the number of misses and false alarms.

The combination of temporal length and box size to yield optimal verification results (that is representative of the “true” model performance) perhaps depends on the storm scale and speed and propagation speed of the storm, but the four cases examined are not conclusive in that regard.

6.4.4 Sensitivity to Microphysics Scheme

In order to investigate model sensitivity to the microphysics scheme, the model was run with the Milbrandt-Yau double-moment 7-class microphysics scheme for the 25 October 2010 event. The Milbrandt-Yau scheme includes separate categories for hail and graupel with double-moment cloud, rain, ice, snow, graupel and hail.

Verification results are tabulated in Table 6.5. On average, the 13-km Thompson scheme performed slightly better compared with Milbrandt-Yau scheme. At 3-km grid resolution, the overall precipitation coverage is better represented by the Thompson scheme, but the Milbrandt-Yau scheme shows significantly better bias than the Thompson scheme in rain areas. In terms of CSI, the two schemes are within the hour-to-hour variation at 3-km grid resolution.

Table 6.5: Verification measurements with Thompson and Milbrandt-Yau (M-Y) microphysics scheme for the 25 October 2010 event. The values are based on a neighborhood box width of 24 (26) km and temporal length of 12 minutes from the model grid resolution of 3 (13) km.

<i>(a) Precipitation coverage</i>				
	Thompson (3 km)	M-Y (3 km)	Thompson (13 km)	M-Y (13 km)
Bias	1.24 (0.18)	0.84 (0.12)	0.94 (0.09)	0.89 (0.10)
POD	0.50 (0.06)	0.39 (0.05)	0.48 (0.04)	0.41 (0.03)
FAR	0.59 (0.05)	0.53 (0.05)	0.48 (0.05)	0.54 (0.04)
CSI	0.29 (0.03)	0.27 (0.03)	0.33 (0.03)	0.28 (0.02)
<i>(b) Snow</i>				
	Thompson (3 km)	M-Y (3 km)	Thompson (13 km)	M-Y (13 km)
Bias	5.23 (1.90)	5.67 (3.13)	5.61 (2.76)	8.98 (7.02)
POD	0.77 (0.29)	0.65 (0.33)	0.88 (0.19)	0.89 (0.16)
FAR	0.84 (0.07)	0.84 (0.11)	0.82 (0.06)	0.86 (0.06)
CSI	0.15 (0.06)	0.15 (0.10)	0.17 (0.06)	0.13 (0.05)
<i>(c) Rain</i>				
	Thompson (3 km)	M-Y (3 km)	Thompson (13 km)	M-Y (13 km)
Bias	1.83 (0.32)	1.15 (0.19)	1.30 (0.20)	1.20 (0.18)
POD	0.61 (0.08)	0.48 (0.05)	0.57 (0.05)	0.47 (0.05)
FAR	0.66 (0.06)	0.57 (0.07)	0.55 (0.07)	0.60 (0.07)
CSI	0.28 (0.04)	0.29 (0.05)	0.33 (0.04)	0.28 (0.04)

Regardless of some differences in the average verification scores between the two schemes, the sensitivity to microphysics scheme is very small in terms of spatial coverage of the three precipitation phases examined. This is in part the 25 October 2010 case being a difficult case to

verify with observations because it included sporadic snow over the complex mountains and transient rain features in the Southeast.

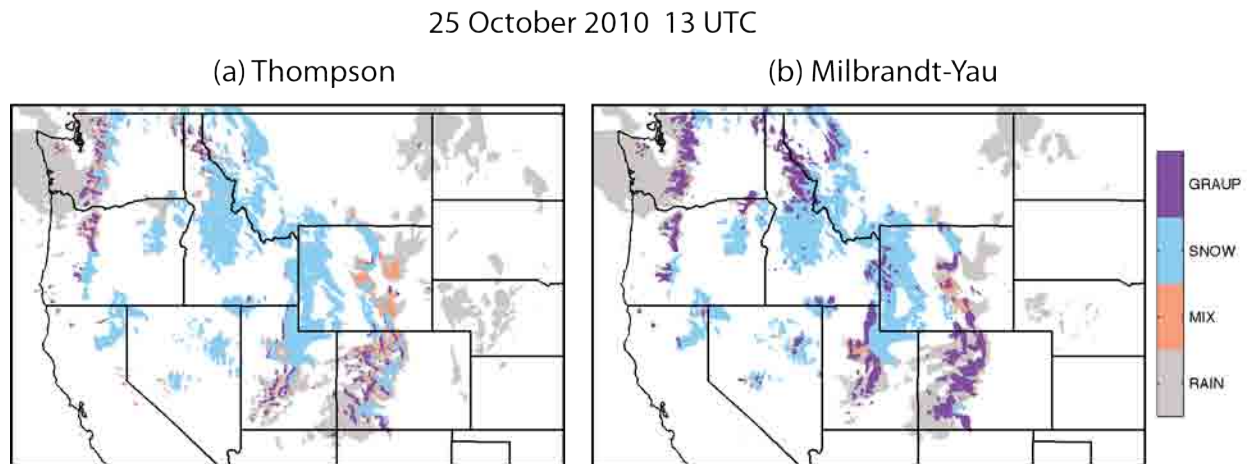


Figure 6.10: Spatial coverage of graupel, snow, mixed-phase precipitation and rain from (a) Thompson and (b) Milbrandt-Yau microphysics schemes valid at 13 UTC on 25 October 2010.

More interestingly, there is a notable difference in the areal coverage of graupel between the two schemes (Fig. 6.10). The Milbrandt-Yau scheme produced significantly larger area of graupel. Figure 6.11 shows one-hour accumulation of total precipitation, snow, and graupel from the two schemes, and hail from Milbrandt-Yau scheme. As illustrated in Fig. 6.11, the 1-hour accumulation amount of graupel by the Milbrandt-Yau scheme is much larger compared to the Thompson scheme; while the snow amount is less than produced by the Thompson scheme. Perhaps, in the Milbrandt-Yau scheme depositional growth of snow and/or collection of cloud droplets by snow, which eventually generate graupel, are much more active. Hail reaching to the ground (Fig. 6.11g) is not realistic in this case (over the particular part of the U.S.). Overall precipitation amount (Figs. 6.11a – b) also suggest some sensitivity. The results are consistent with a recent study by Liu et al. (2011) that showed a significant difference in precipitation amount among various microphysics schemes based on a WRF simulation of cool-season precipitation over the Colorado Rockies. In their study, the Thompson microphysics scheme outperformed other schemes compared to observations.

25 October 2010 13 UTC

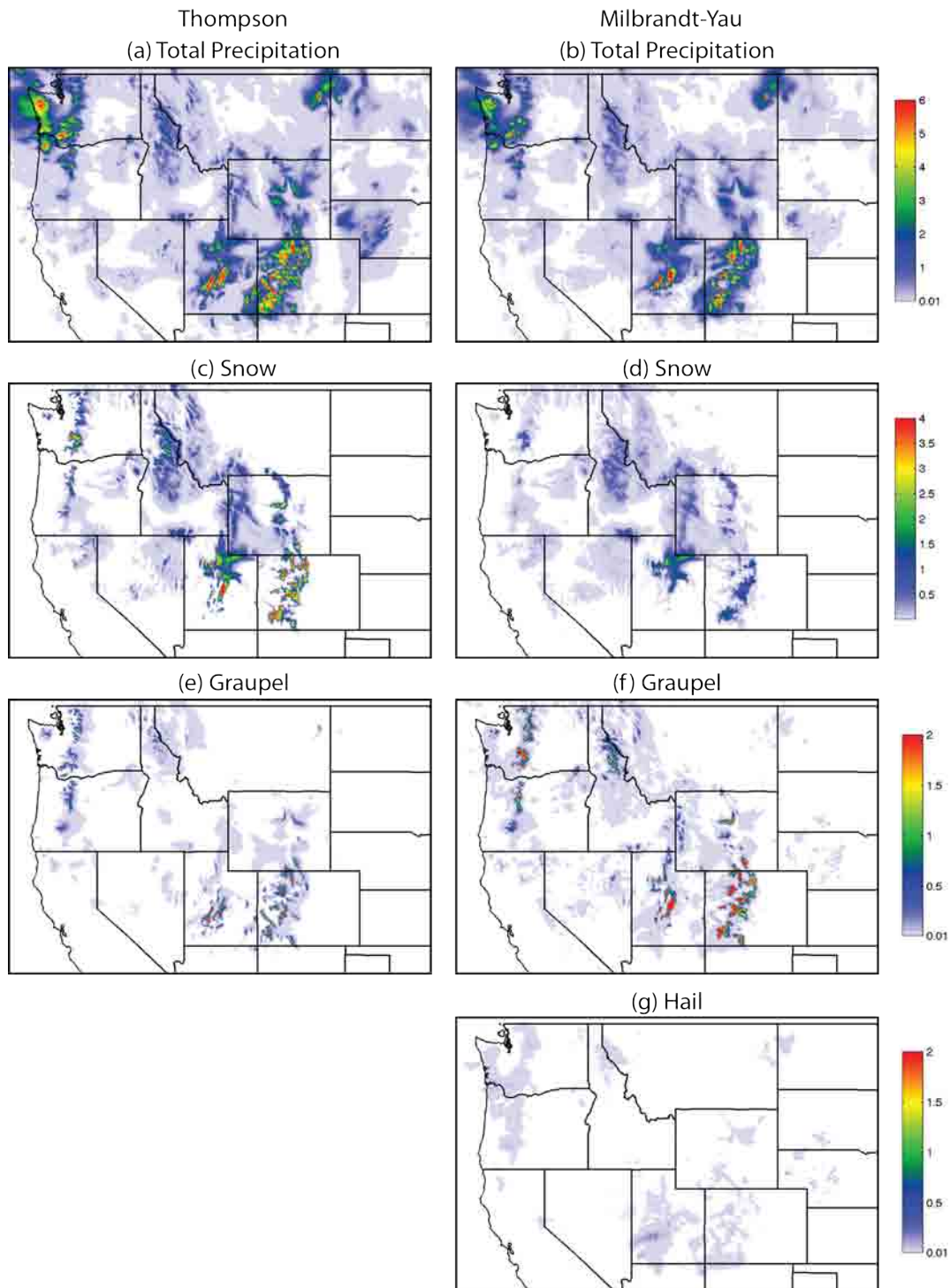


Figure 6.11: One-hour accumulation of (a)-(b) total precipitation, (c)-(d) snow, (e)-(f) graupel, and (g) hail simulated with Thompson (left panels) and Milbrandt-Yau (right panels) schemes. Values are in millimeters. The 1-hour accumulation is valid at 13 UTC on 25 October 2010.

6.5 Key Remarks and Recommendations

Forecast skill in predicting areal coverage of cool-season precipitation and phase was evaluated using ASOS observation data. Key results from the current study are as follows:

- Examination of various masks on model precipitation field showed that representing areas of rain, snow, and mixed-phase works well when a mass (amount)-based threshold depending on rain-snow and rain-graupel density ratios are used to mask precipitation field. For the cases examined, a mask level of 1×10^{-3} mm/time-step for rain, 1×10^{-4} for snow, and 4×10^{-4} for graupel showed the best agreement with the observations. Using a constant threshold value of $< 1 \times 10^{-4}$ mm/time-step for all hydrometeors generated too many false alarms in rain regions; using a constant value of 1×10^{-3} resulted in too many miss in snow areas.
- For strong synoptically forced frontal systems over the relatively flat central U.S., the model performed well. In particular, snow, rain, and mixed-phase coverage agreed well with surface observations. In addition, the rain-snow transition was well represented by the model.
- For relatively weaker systems, which produced scattered precipitation and narrow rain-snow transition boundaries, the skill scores were significantly lower compared to strong precipitating systems. The data indicate that precipitation occurs generally in the correct region at correct times. However, comparisons with observations are complicated for such precipitation because any location and/or temporal offsets (where and when the precipitation occurred) affect the skill scores.
- The rain-snow boundary over complex terrain, particularly in the western U.S., is difficult to forecast. Verifying the model data is also challenging for the Intermountain west because surface observation sites are sparsely distributed and local terrain induced effects may play a significant role in the local distribution of precipitation.
- The 13-km resolution model produced significantly larger bias in the rain regions compared with the 3-km model due to over-forecasts of cumulus-scheme based convections. The skill difference between the 3- and 13-km resolution model runs was small in the snow areas.
- In terms of precipitation phase on the ground, there was no significant sensitivity to the microphysics schemes for the 25 October 2010 case. However, there was a notable difference in snow, rain, graupel amounts. This calls for comparison of the hourly totals to surface observations because an accurate prediction of hourly accumulation of particular precipitation types is essential for aviation purposes.

Future work and recommendations for model enhancement are:

- Apply hydrometeor density based thresholds to precipitation fields for representing areal coverage of snow, rain, graupel, and mixed-phase precipitation.
- Evaluate precipitation type and amount in the Western U.S. with other sources of surface verification data (e.g., SNOTEL).

- Examine if the model has a time-lag in predicting the snow-rain transition using additional winter cases.

References for Section6

- Liu, C., K. Ikeda, R. Rasmussen, G. Thompson, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado Headwaters region: sensitivity to physics parameterizations. *Monthly Weather Review*, doi: 10.1175/MWR-D-11-00009.1, in press.
- Grell, G. A. and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, **29**, No 14., 10.1029/2002GL015311, 2002.

7. Summary of Findings and Outlook

This report documented sensitivity experiments geared toward elucidating the potential impacts of model configuration changes on the prediction of aviation weather hazards. This year's effort was focused on aspects of convection, turbulence, in-flight icing, and precipitation type. Model simulation experiments emulated the WRF Rapid Refresh (WRF-RR) and High-Resolution Rapid Refresh (HRRR) models operating at 13 km and 3 km spatial resolution, respectively, using the latest versions of the WRF code and configuration files as used by NOAA/ESRL. These experiments were focused on seven high-impact days from 2010 and 2011, covering both winter and summer cases that exhibited convective storms, turbulence, and icing conditions. Forecast-sensitivity experiments addressed horizontal and vertical resolution, microphysics and boundary layer physics, and land surface parameterization schemes, and the diffusion coefficient. In addition, an assessment of the HRRR operationally run during the summer 2011 was performed, especially looking at the model's skill in predicting mesoscale convective storm systems. Performance assessment was measured against surface (in situ and remotely sensed) and upper air observations, and PIREPs, including regional differences.

These analyses focused on sensitivity impacts of numerical weather prediction model configurations on a variety of aviation hazards need to be extended. The analyses presented in this report are too limited to provide conclusive evidence and thus clear guidance for improving the 3-km HRRR and 13-km WRF-RR model configurations.

7.1 Convection

Riding on the recent developments in WRF and high resolution data assimilation techniques, the 3 km HRRR forecasts did reasonably well on convective systems of relatively large spatial and temporal extent, such as lines of convection, regional scattered convection, and winter storms. It usually places the systems in roughly the same region as observed, depicts storm structures with realism, and predicts the convective mode correctly. On the other hand, errors are evident in the predicted mesoscale details, such as in the exact location and orientation of line storms and storm clusters, initiation and maintenance of MCSs, and timing and strength of sea breeze convection. Such errors can be significant even in very short-term, say 1 – 3 hour forecasts.

Generally speaking, the precipitation forecasting skills for cold season cases are significantly better than those for warm season cases, for both the 3 km and 13 km runs. However, a problem of over-prediction was identified with the 13 km simulations. The model at 13 km resolution also seems to experience a “spin-down” period in the first hour. Large areas of false-alarm precipitation are predicted immediately following the initialization, especially over the ocean in the southeast part of the domain. An explanation may be that since the RUC tends to under-forecast convection, the RUC analyses that are tuned up for RUC may be too hot for starting WRF forecasts.

Quantitative skills of the 3 km baseline simulations show cold and wet biases in surface temperature and moisture, and over-forecasts of surface wind speed. The errors in temperature and moisture gradually grow with forecast length and correlated with the diurnal cycle. The wind speed errors grow rapidly in the first forecast hour, and the wind direction errors (RMSE) are the largest in the initial conditions. This may be an indication of small-scale mismatches and imbalances in the initial conditions. Surface observation also reveal that the simulated

temperature gradients across sharp boundaries tend to be smoothed. Verification using observed soundings and surface observations revealed obvious model deficiencies at low levels in terms of wind direction, temperature, and humidity. Detailed studies on how to improve the analysis and forecast of low-level representations in models are key to improve short-term convective forecasting.

Sensitivity experiments show that, using different physics options usually does not adversely change the forecast of convection. However, prediction of the details of storms can be significantly impacted. Moderate sensitivity is seen in rainfall forecasts to PBL and microphysics schemes. On the other hand, it is not clear which scheme has better skills for convective forecasting over the entire CONUS. The skills of the physics packages may be regime dependent. It is also found that the physics option may affect certain model fields (e.g., microphysics scheme affects hydrometeor partitions) more than they do to rainfall. The YSU PBL scheme is found to severely over-estimate PBL height and therefore PBL mixing. Two modifications to the standard YSU were therefore proposed (Yubao Liu and Greg Thompson of NCAR, personal communications) and were found to mitigate the problems to some degree.

Post-season analysis found that the operational HRRR forecasts missed more than 30% of the nightly MCS initiation during June-August 2010. To better resolve the forcing of and therefore to improve the forecast of nocturnal MCSs will continue to be a challenge for the HRRR model.

For the next year, we plan to further analyze the baseline and sensitivity simulations performed in FY11. Attention will be given to the evaluation of simulations focused on regional domains and individual convective systems. Additional sensitivity experiments will be conducted with regard to the model initial conditions and physics schemes. These experiments and analyses will also focus on a particular forecast challenge, such as the initiation and up-scale growth of MCS on 19 July 2010. Previous experiments have shown that model initial conditions are still the controlling factor in convective forecast skills. The verifications against observations in this study provided further evidence that the current model has significant problems with the boundary layer, especially for wind and temperature; both are sensitive fields for convective initiation. We believe that the key for improving the analyses and, therefore, the forecasts of these fields and rainfall, are to improve the surface data and radar data assimilation in concert with improving the boundary layer physics.

7.2 Turbulence

The analyses focused on turbulence aspects resulted in the following findings:

- Different model configurations have different affects on different diagnostics;
- Largest affect is with 3 km horizontal grid;
- Doubling vertical resolution everywhere makes little difference;
- Diff = 0 makes little difference;
- YSU PBL makes little difference except it provides no upper level SGSTKE;
- Limited statistical evaluations indicate the overall forecast skill is no better for the 3 km grid than for the 13 km grid;
- The MYJ turbulence parameterization performs very poorly compared to post-processed turbulence diagnostics for both the 3-km and 13-km grids.

It is not clear how to explain these results, especially the observation that higher resolutions, both

horizontal and vertical, do not lead to better results, either in the case studies or the statistical analyses. This might be in part due to the cases chosen, which were primarily driven by convection considerations. Due to the highly transient nature of convectively-induced turbulence, these cases would be more difficult to handle from the turbulence perspective. Perhaps future work should concentrate on cases that are more of the CAT variety. Also, this study did not consider mid- or upper-levels. It could be that higher resolution is more beneficial for turbulence sources active in this part of the atmosphere, but this also needs to be investigated. Finally, it is not clear why the MYJ scheme does not provide better statistical performance. This needs to be investigated further as well.

7.3 In-flight Icing

The 3-km WRF was able to generate higher (positive and negative) vertical velocities than the 13-km WRF. This occurred more often in the summer cases. The stronger updrafts in the 3-km WRF generated additional condensate compared to the 13-km WRF.

While both models did an excellent job predicting a model cloud (either ice or liquid) at the PIREP locations, they appeared to over predict ice and under-predict supercooled liquid water. This is based on extent of coverage, not quantity of condensate.

Neither model correctly identified icing conditions at temperatures colder than -12°C. They both correctly identified cloud (ice or liquid) but incorrectly predicted ice rather than supercooled liquid water the majority of the time.

Wave-like structures were observed across the 3-km domain. They appear to be associated with the predictions of condensate. These waves may be model artifacts or they may be real structures. If they are deemed significant in terms of forecasting aviation weather hazards, they should be investigated further.

7.4 Precipitation Type

Forecast skill in predicting areal coverage of cool-season precipitation and phase was evaluated using ASOS observation data. Key results from the current study are as follows:

- Examination of various masks on model precipitation field showed that representing areas of rain, snow, and mixed-phase works well when a mass (amount)-based threshold depending on rain-snow and rain-graupel density ratios are used to mask precipitation field. For the cases examined, a mask level of 1×10^{-3} mm/time-step for rain, 1×10^{-4} for snow, and 4×10^{-4} for graupel showed the best agreement with the observations. Using a constant threshold value of $< 1 \times 10^{-4}$ mm/time-step for all hydrometeors generated too many false alarms in rain regions; using a constant value of 1×10^{-3} resulted in too many miss in snow areas. Constant threshold value for all precipitation types.
- For strong synoptically forced frontal systems over the relatively flat central U.S., the model performed well. In particular, snow, rain, and mixed-phase coverage agreed well with surface observations. In addition, the rain-snow transition was well represented by the model.
- For relatively weaker systems, which produced scattered precipitation and narrow rain-snow transition boundaries, the skill scores were significantly lower compared to strong

precipitating systems. The data indicate that precipitation occurs generally in the correct region at correct times. However, comparisons with observations are complicated for such precipitation because any location and/or temporal offsets (where and when the precipitation occurred) affect the skill scores.

- The rain-snow boundary over complex terrain, particularly in the western U.S., is difficult to forecast. Verifying the model data is also challenging for the Intermountain west because surface observation sites are sparsely distributed and local terrain induced effects may play a significant role in the local distribution of precipitation.
- The 13-km resolution model produced significantly larger bias in the rain regions compared with the 3-km model due to over-forecasts of cumulus-schemed based convections. The skill difference between the 3- and 13-km resolution model runs was small in the snow areas.
- In terms of precipitation phase on the ground, there was no significant sensitivity to the microphysics schemes for the 25 October 2010 case. However, there was a notable difference in snow, rain, graupel amounts. This calls for comparison of the hourly totals to surface observations because an accurate prediction of hourly accumulation of particular precipitation types is essential for aviation purposes.

Future work and recommendations for model enhancement are:

- Apply hydrometeor density based thresholds to precipitation fields for representing areal coverage of snow, rain, graupel, and mixed-phase precipitation.
- Evaluate precipitation type and amount in the Western U.S. with other sources of surface verification data (e.g., SNOTEL).
- Examine if the model has a time-lag in predicting the snow-rain transition using additional winter cases.